

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA



TESIS DE GRADO

REDES NEURONALES
APLICADOS A
PROCESOS ESTOCÁSTICOS

PARA OPTAR AL TÍTULO DE LICENCIATURA EN INFORMÁTICA
MENCION INGENIERÍA DE SISTEMAS INFORMÁTICOS

AUTOR : MIGUEL DE SEVILLA CHAVEZ GORDILLO
TUTORA : Lic. ROSA FLORES MORALES M. Sc.
REVISORA : Lic. CARMEN ROSA HUANCA QUISBERT
ASESOR : Lic. JUAN CARLOS FLORES LÓPEZ

La Paz - Bolivia
2006

Dedicatoria

*Dedicado a mí mama **Juana Gordillo***

A mis hermanos Errol, Alejandro, Miguel, Ismael, Ariel, Esther, Zulema, Yola por toda la alegría que siempre compartimos en el transcurso de nuestras vidas.

Agradecimientos

A ti mamá que siempre me apoyaste en todo lo que me tocó vivir y haberme permitido cultivar principios nobles que hoy rigen mi vida, depositaste tu confianza en mí y es a ti a quien debo el logro de esta meta.

A mis hermanos Errol, Alejandro, Miguel, Ismael, Ariel, Esther, Zulema, Yola por acompañarme en todo el recorrido de nuestras vidas.

Agradecer a la Licenciada Rosa Flores Morales M. Sc. Tutora, quien me guió por el sendero correcto en el proceso de investigación y que hoy me permite concluir la presente Tesis.

A la Licenciada Carmen Rosa Huanca Quisbert Docente revisora, quien hizo observaciones útiles permitiendo hacer mejoras tras cada revisión.

Al Licenciado Juan Carlos Flores Docente de la carrera de Estadística por haberme encaminado en el área de series temporales y su aplicabilidad, por toda la colaboración, tiempo y dedicación puesta en el presente trabajo.

A mis amigos Ximena, a Fernando que con comprensión y perseverancia contribuye a forjar el conocimiento administrando la Biblioteca... Gracias por su colaboración desde nuestro inicio en la carrera hasta hace algunos días que cerramos una etapa más en nuestras vidas.

Resumen

El análisis de series temporales ha pasado a ocupar un lugar central en el estudio del comportamiento de las variables correspondientes a eventos: Biológicos, Medio ambiente, Finanzas, Empresariales, Medicina, los cuales permitirán planificar con un mayor grado de confiabilidad las actividades relativas a dichos eventos.

En las últimas décadas hemos podido asistir a la aparición de modelos y métodos estadísticos de gran importancia tales como los modelos $ARIMA(p, d, q)$, modelos de transferencia y los modelos de regresión dinámicos de Koyck y Almon, dichos modelos están orientados al trabajo con datos provenientes de series temporales.

El presente trabajo esta orientado a incorporar el modelo de redes neuronales artificiales como una herramienta más para el estudio de datos provenientes de series temporales.

La idea consiste en asociar una serie temporal a un conjunto de entrenamiento y valores deseados de forma que la red neuronal aprenda a reconocer esta sucesión temporal mediante un proceso de entrenamiento.

Se disponen de n -observaciones, si bien se pueden asociar $(n-1)$ observaciones para el vector de entrenamiento $(y_1, y_2, \dots, y_{n-1})$, en el proceso de pruebas se verificó que tres observaciones antecedentes $(y_{i-3}, y_{i-2}, y_{i-1})$ a una consecuenta y_i es un criterio óptimo para que la red neuronal pueda ajustar la serie temporal con un error cuadrático mínimo.

Se dispone del criterio de asociación de la infamación, pero no obstante la arquitectura de la red neuronal para toda sucesión temporal es independiente, se deben realizar una serie de entrenamientos hasta encontrar el error cuadrático mínimo, entonces dicha arquitectura de red corresponderá a la serie estudiada.

Temáticas de investigación relacionadas con el presente trabajo son relativas a: redes neuronales, procesos estocásticos y series temporales.

Noviembre 2006.

Índice general

1. Introducción	7
1.1. Planteamiento del problema	8
1.2. Objetivos	15
1.2.1. Objetivo general	15
1.2.2. Objetivos específicos	15
1.3. Hipótesis	15
1.4. Alcances	15
1.5. Aportes	16
2. Redes Neuronales	17
2.1. Introducción	17
2.2. La neurona	18
2.2.1. Criterio de activación	19
2.2.2. Funciones de transferencia	19
2.3. Topología de las redes neuronales	21
2.3.1. Redes monocapa	22
2.3.2. Redes multicapa	22
2.3.3. Consideraciones de la red multicapa	23
2.4. Interpretación geométrica	25
2.4.1. La dimensión $n+1$	28
2.5. Algoritmos de aprendizaje	30
2.5.1. Método del gradiente descendente	30
2.5.2. Procedimiento Widrow-Hoff	33
2.5.3. Procedimiento delta generalizado	35
2.6. Metodo Backpropagation o retropropagación	37
2.6.1. Análisis de " δ'_i " en la última capa	40
2.6.2. Análisis de " δ''_i " en las capas intermedias	41
3. Series Temporales	46
3.1. Introducción	46
3.2. Proceso estocástico	46
3.2.1. Ruido blanco	48
3.2.2. Proceso estocástico estacionario	48
3.3. Modelos autorregresivos AR	50
3.3.1. Modelo AR(1)	50

3.3.2.	Modelo AR(2)	53
3.3.3.	Modelo AR(p)	56
3.4.	Modelos de medias móviles MA	60
3.4.1.	Modelo MA(1)	60
3.4.2.	Modelo MA(2)	63
3.4.3.	Modelo MA(q)	66
3.5.	Modelos mixtos autorregresivos-medias móviles ARMA	68
3.5.1.	Modelo ARMA(1,1)	69
3.5.2.	Modelo ARMA(p,q)	73
3.6.	Modelo ARIMA	75
3.6.1.	Identificación del modelo ARIMA	78
3.6.2.	Instrumentos estadísticos de identificación	79
4.	Construcción de la Red Neuronal	84
4.1.	Introducción	84
4.2.	Muestra aleatoria	84
4.3.	Conjunto de entrenamiento	85
4.4.	Dimensión de los vectores de entrenamiento	87
4.5.	Diseño experimental	90
4.5.1.	Identificación de variables de la hipótesis	90
4.5.2.	Instrumento de medida	90
4.5.3.	Registro experimental	91
4.6.	Prototipo	92
4.6.1.	Descripción del prototipo	92
4.6.2.	Algoritmos	94
5.	Resultados Prueba Experimental	96
5.1.	Datos para el entrenamiento	96
5.2.	Pruebas	97
5.3.	Resultados	101
5.4.	Pesos para índice de exportaciones	106
5.5.	Comparación de resultados	107
5.6.	Discusión de resultados	107
5.7.	Aplicación en Medicina	109
6.	Conclusiones y Recomendaciones	110
6.1.	Conclusiones	110
6.2.	Recomendaciones	112

Capítulo 1

Introducción

En el presente trabajo se aplicó los modelos ARIMA y modelos de redes neuronales artificiales a diferentes series temporales, con el objetivo de obtener modelos generadores que permitirán realizar procesos de ajuste y predicción de las series estudiadas.

A continuación se expone su estructura

El capítulo 1, contiene la temática de investigación, se expone el problema a abordar orientando los objetivos a la solución de dicho problema.

EL capítulo 2, orientado al estudio de las redes neuronales artificiales, se expone el procedimiento delta generalizado que es la base fundamental del algoritmo *“Backpropagation”*.

El capítulo 3, proporciona una panorámica de los procesos estocásticos, los modelos $AR(p)$, $MA(q)$, $ARMA(p, q)$ base del modelo $ARIMA(p, d, q)$.

El capítulo 4, provee los fundamentos que permiten la obtención del conjunto de entrenamiento y valores deseados de la red neuronal.

El capítulo 5, se obtiene la topología de red neuronal óptima para las series estudiadas. Se contrasta los resultados obtenidos entre los modelos ARIMA y redes neuronales.

El capítulo 6, esta orientado a recopilar todas las conclusiones y recomendaciones obtenidas a partir de la fase experimental.

1.1. Planteamiento del problema

La teoría propuesta por los estadísticos Box y Jenkins (1970) permite asociar un comportamiento estacionario a un único modelo expresado en términos autorregresivos AR(p), medias móviles MA(q) y autorregresivos medias móviles ARMA(p, q). Para el caso no estacionario se asocia a un modelo ARIMA(p, d, q) esta relación se observa en las figuras 1.1, 1.2, 1.3, 1.4

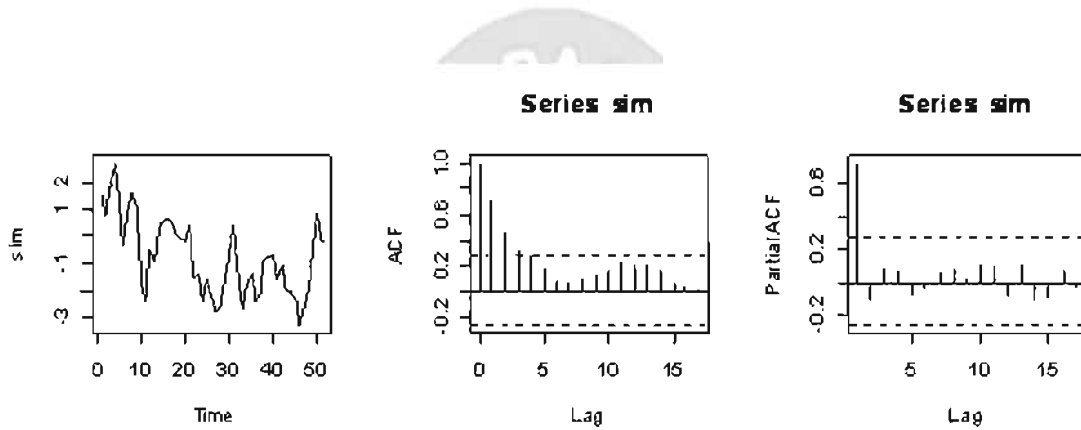


Figura 1.1: Modelo autorregresivo AR(1) $\phi_1 = 0,7$ $y_t = \phi_1 y_{t-1} + \varepsilon_t$
 Fuente: Elaboración propia.

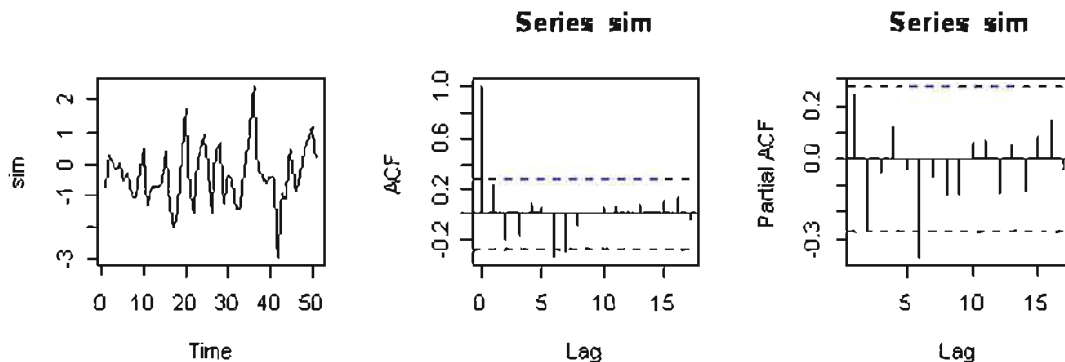


Figura 1.2: Modelo de medias móviles MA(1) $\theta_1 = 0,3$ $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$
 Fuente: Elaboración propia.

En la vida real los eventos: Biológicos, Medio ambiente, Finanzas, Empresariales, Medicina; No presentan estos fascinantes comportamientos estacionarios, al contrario en ocasiones pueden ser bastante irregulares y caóticos.

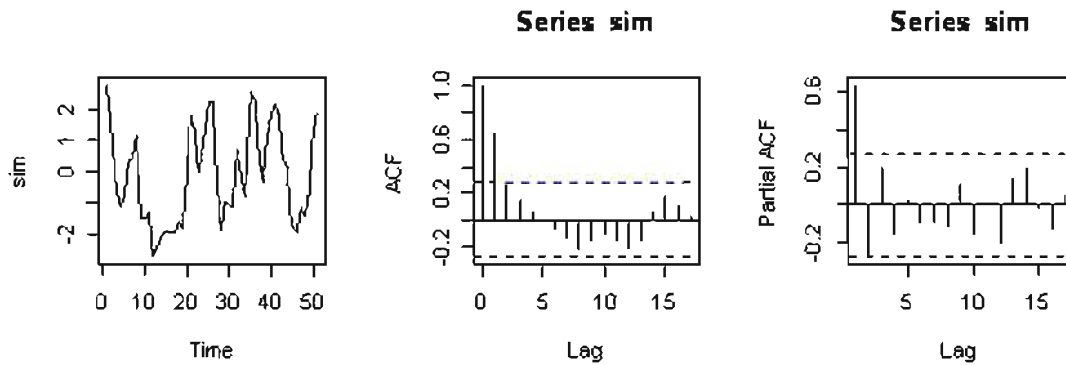


Figura 1.3: Modelo autorregresivo media móvil ARMA(1,1) $\phi_1 = 0,7$ $\theta_1 = 0,3$ $y_t = \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$

Fuente: Elaboración propia.

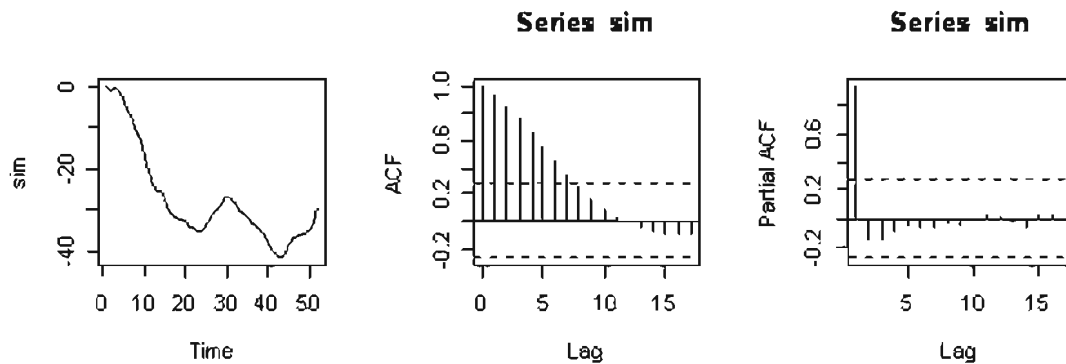


Figura 1.4: Modelo autorregresivo media móvil integrado ARIMA(1,1,1) no estacionario $\phi_1 = 0,7$ $\theta_1 = 0,3$ $(1 - \phi_1 L)(1 - L)^{d=1} y_t = (1 - \theta_1 L) \varepsilon_t$

Fuente: Elaboración propia.

El proceso de hallar el modelo generador de la serie temporal que permita realizar predicciones e interpolación de resultados no es tarea trivial requiere de un *Analista de Series Temporales*, el proceso consiste en:

- a) Se debe observar si la serie estudiada es o no estacionaria, si no lo fuere se transforma a un serie estacionaria mediante un proceso de integración que consiste en aplicar diferencias entre las observaciones sucesivas, generalmente son de primer y segundo orden para convertirla a una serie estacionaria.
- b) Con la serie estacionarizada se hace un estudio de las funciones de autocorrelación FAC y FACP, se deben verificar los comportamientos del cuadro 1.1.

MODELO	FAC	FACP
AR(p)	Decrecimiento rápido de tipo geométrico puro y geométrico con alternancia de signos, sinusoidal o mezcla de varios tipos	Se anula para retardos superiores a p
MA(q)	Se anula para retardos superiores a q	decrecimiento rápido de tipo exponencial y/o sinusoidal.
ARMA(p,q)	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas

Cuadro 1.1: FAC y FACP en los modelos AR(p),MA(q),ARMA(p,q)

Fuente: Uriel E. Peiro A.

- c) De acuerdo al comportamiento de las funciones de autocorrelación FAC, FACP se identifica el modelo generador AR(p), MA(q), ARMA(p, q) y los factores de retardo p, q de la serie temporal en estudio.
- d) Con el modelo teórico asociado es necesario realizar pruebas que sustenten la calidad del modelo obtenido, mediante los criterios de información Akaike AIC y BIC Bayesian Information Criterion, se elige aquel que minimiza el AIC y BIC.

En el proceso de identificación del modelo generador de la serie temporal quizá no exista ambigüedad en las funciones FAC y FACP para los modelo AR(p), MA(q) aunque el factor de retardo puede estar sujeto a la apreciación del analista.

Cuando el proceso no corresponde a un modelo estacionario AR(p) o MA(q), las funciones de autocorrelación se comportan de la siguiente manera, los primeros valores iniciales no tienen un patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas. Se puede conocer el grado de integración o diferencia pero no se precisa con certeza los factores de retardo p, q que generalmente están restringidos al conjunto $\{0, 1, 2\}$, a fin de conocer el modelo generador se deben realizar varios procesos ARIMA(p, d, q).

Finalmente de todos los modelo obtenidos se tiene la relación de los criterios de información AIC y BIC cuyo valor mínimo corresponderá al modelo óptimo el cual permi-

tirá realizar tareas de predicción e interpolación de datos.

Sustento del problema

Se trabajara con la serie temporal *Índice de Precios al Consumidor IPC* , que puede observarse en la figura 1.5, los datos fueron obtenidos del *Instituto Nacional de Estadística INE* entre los periodos jun-00 sep-05.

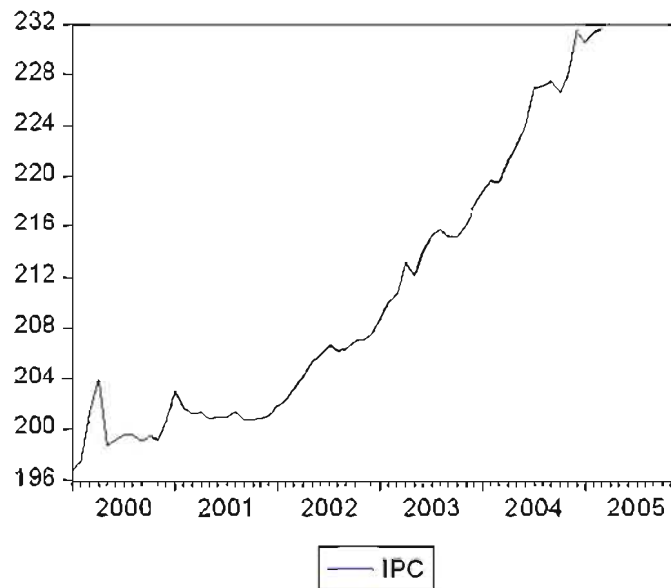


Figura 1.5: Comportamiento del IPC
Fuente: Elaboración propia.

Aplicando el proceso de análisis de series temporales se observa que no es un proceso estacionario, por lo que se aplica un proceso de diferencias para convertirla en serie estacionaria, se evalúan las funciones FAC, FACP que proveen los siguientes resultados observables en la figura 1.6

El cuadro del extremo izquierdo en la figura 1.6 corresponde a un proceso de integración de orden dos, análogamente el de la derecha de orden uno, empezaremos desarrollando la búsqueda del modelo para un orden dos en el cual se observa que los rezagos tienden a cero a partir de $p=4$, entonces para un grado de diferencia igual a dos el modelo tiende a un autorregresivo AR(4).

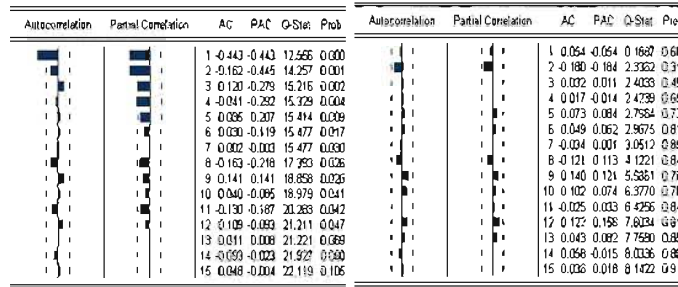


Figura 1.6: Funciones de autocorrelación FAC y autocorrelación parcial FACP
Fuente: Elaboración propia.

Se estiman los parámetros del modelo al igual que los criterios AIC, BIC y se realiza el ajuste, puede observarse ello en la figura 1.7

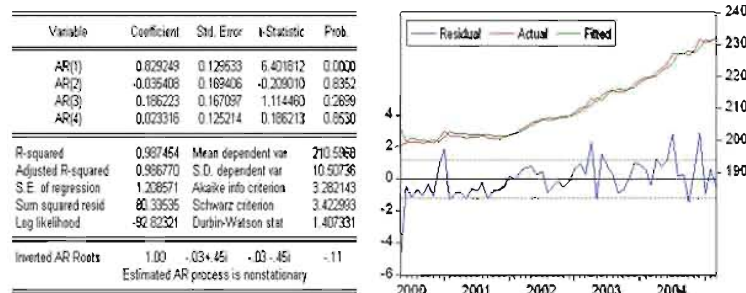


Figura 1.7: Parámetros estimados y ajuste del modelo AR(4)
Fuente: Elaboración propia.

El modelo obtenido es un posible candidato del buscado ya que los criterios de información Akaike y Bayesiano son razonablemente mínimos.

En la figura 1.6 el cuadro derecho que corresponde a un proceso de integración de orden uno, las funciones de auto correlación no tienen un patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas razón por el cual queda claro que corresponde a un modelo ARMA(p, q), pero no se puede precisar con certeza los valores que tomaran los factores de retardo p, q

Debido a que son series no estacionarias, las funciones FAC, FACP sugieren analizar los modelos ARIMA(1, 1, 1) ARIMA(2, 1, 2) ARIMA(2, 1, 1) ARIMA(1, 1, 2).

Aplicando el proceso ARIMA se obtiene los siguientes resultados, en la figura 1.8 se observa el modelo ARIMA(1,1,1), en la figura 1.9 al modelo ARIMA(2,1,2), en la figura 1.10 al modelo ARIMA(2,1,1), en la figura 1.11 al modelo ARIMA(1,1,2).

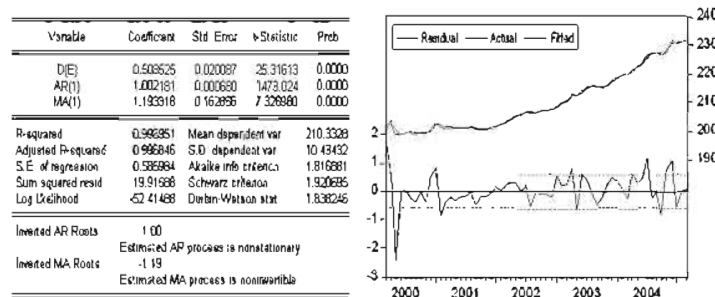


Figura 1.8: Parámetros estimados y ajuste del modelo ARIMA(1,1,1)

Fuente: Elaboración propia.

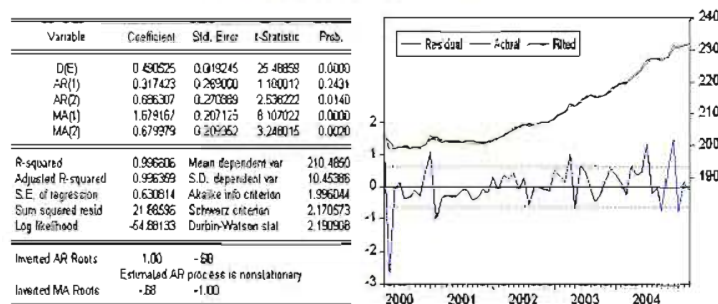


Figura 1.9: Parámetros estimados y ajuste del modelo ARIMA(2,1,2)

Fuente: Elaboración propia.

Los criterios de información AIC, BIC de los diferentes modelos ARIMA(p,d,q) los cuales se observan en el cuadro 1.2 proveerán información acerca del modelo óptimo.

El modelo óptimo es *ARIMA(1,1,1)* debido a que los criterios de información AIC, BIC son lo mínimos de entre todos en el cuadro 1.2

El problema en términos concretos es: *La búsqueda del proceso generador ARIMA*, que depende de la apreciación del analista de series temporales donde jugará un papel fundamental su habilidad y destreza ligado a los factores humanos.

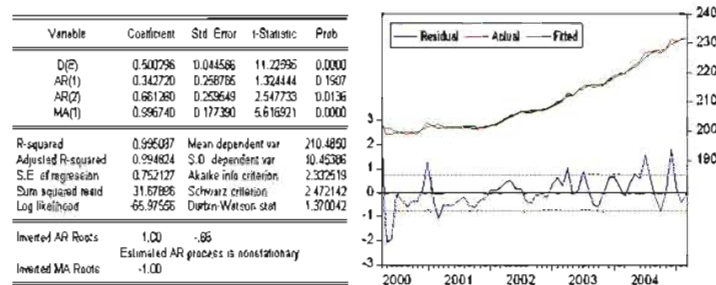


Figura 1.10: Parámetros estimados y ajuste del modelo ARIMA(2,1,1)

Fuente: Elaboración propia.

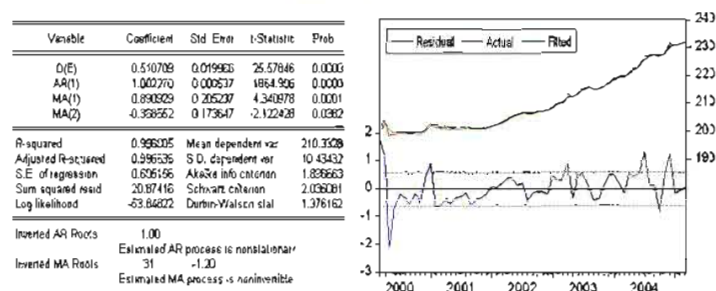


Figura 1.11: Parámetros estimados y ajuste del modelo ARIMA(1,1,2)

Fuente: Elaboración propia.

El presente trabajo de investigación formulo aplicar la red neuronal artificial de alimentación hacia adelante denominada “backpropagation” a fin de hallar el modelo generador ARIMA.

Modelo ARIMA	AIC	BIC
AR(4)	3.282143	3.422993
ARIMA(1,1,1)	1.816881	1.920695
ARIMA(2,1,2)	1.996044	2.170573
ARIMA(2,1,1)	2.332519	2.472142
ARIMA(1,1,2)	1.896663	2.035081
Mínimo	1.816881	1.920695

Cuadro 1.2: Criterios de información AIC, BIC

Fuente: Elaboración propia

1.2. Objetivos

1.2.1. Objetivo general

*Modelar un proceso estocástico mediante redes neuronales artificiales “Backpropagation” de alimentación hacia delante denominadas **feedforward**.*

1.2.2. Objetivos específicos

- Establecer criterios respecto a la asociación y representación de la información de entrada y salida en la red neuronal artificial.
- Formular la topología adecuada de la red neuronal que será aplicada a un proceso estocástico.
- Contrastar los resultados obtenidos mediante redes neuronales artificiales y modelos ARIMA.
- Desarrollar el prototipo que permita obtener el modelo generador de un proceso estocástico.

1.3. Hipótesis

La red neuronal artificial “*Backpropagation*” obtiene un modelo generador de un proceso estocástico con un 95% de confiabilidad en comparación con el modelo ARIMA.

1.4. Alcances

La investigación comprende un estudio de las siguientes áreas a incursionar:

- Desarrollar el marco teórico relativo a las redes neuronales artificiales, profundizando en el estudio de las redes multicapa cuya base se debe al procedimiento delta generalizado que permitió generar el algoritmo para el entrenamiento de redes con arquitectura multicapa.

- Desarrollar el marco teórico respecto a los modelos estadísticos: $AR(p)$ modelo autorregresivo, $MA(q)$ modelo de medias móviles, $ARMA(p, q)$ modelo autorregresivo y medias-móviles, $ARIMA(p, d, q)$ modelo autorregresivo y medias-móviles integrado.
- Deducir criterios mediante los cuales una sucesión de variables aleatorias temporales y_{t_1}, \dots, y_{t_n} serán transformados en un conjunto de entrenamiento de la red neuronal artificial.
- Contrastar los resultados obtenidos por los modelos estadísticos $ARIMA(p, d, q)$ y los modelos obtenidos mediante redes neuronales artificiales.
- Desarrollar el prototipo que permita aplicar la red neuronal artificial a los procesos estocásticos, el software se desarrollara en Matlab 7.0 contendrá procesos de normalización de datos, fase de entrenamiento, modulo para la obtención del modelo generador.

1.5. Aportes

- Formulación optima de la arquitectura de la red neuronal artificial aplicada a un proceso estocástico.
- Criterios formales mediante los cuales una sucesión de variables aleatorias temporales, son transformados en un conjunto de entrenamiento de la red neuronal artificial.
- Elaboración del prototipo que permitirá crear el modelo generador de un proceso estocástico.

Capítulo 2

Redes Neuronales

2.1. Introducción

Las redes neuronales son estructuras con la propiedad de aprendizaje a partir de ejemplos, son utilizados en diferentes campos de aplicación tales como: visión artificial, reconocimiento del habla, escritura, robótica, optimización.

Los tres elementos clave de los sistemas biológicos que pretende simular son: el procesamiento en paralelo, la memoria distribuida y la adaptabilidad.

Definición de red neuronal

Las redes neuronales artificiales son redes interconectadas masivamente en paralelo son adaptativas y con organización jerárquica, interactúan con el mundo real de forma similar a la neurona biológica.

En cuanto a sus características relevantes tienen la capacidad de aprender mediante una fase de entrenamiento (o experiencia), pueden generalizar ciertas propiedades a partir de casos específicos.

Es un modelo matemático que representa al funcionamiento de la neurona biológica, las entradas son vectores x_i de dimensión n asociados a coeficientes w_i denominados pesos los cuales forman la ecuación $w_1x_1, \dots, w_nx_n = \theta$ que debe rebasar el umbral θ .

2.2. La neurona

Las neuronas son estructuras básicas de la red neuronal, constan de un vector de entradas $X = (x_1, \dots, x_n)$ asociados a un conjunto de pesos $W = (w_1, \dots, w_n)$ correspondientes a la misma dimensión n , se requiere de un parámetro denominado *bias* " θ " el cual permite desplazar el hiperplano respecto al origen en una base ortonormal R^n . Se observa la representación de la neurona en la siguiente figura 2.1.

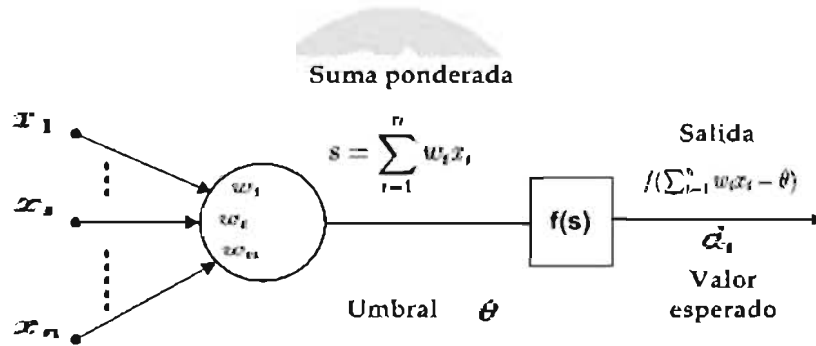


Figura 2.1: Estructura de la neurona

Fuente: Nils, J. Nilsson

Se denomina hiperplano a la expresión algebraica que cumple la siguiente propiedad

$$w_1 x_1 + \dots, w_i x_i + \dots + w_n x_n = \theta \quad (2.1)$$

La idea consiste en evaluar la ecuación bajo el siguiente criterio

$$s = w_1 x_1 + \dots + w_i x_i + \dots + w_n x_n - \theta \quad (2.2)$$

Al que se denomina suma ponderada o producto escalar, el que proveerá valor cero si el punto $X = (x_1, \dots, x_n)$ esta exactamente en el hiperplano, análogamente proveerá valores que permitan situarlos en un extremo u otro del hiperplano, para ello se requieren criterios de decisión representado por la denominada "*función de activación*" cuya propiedad es permitir clasificar correctamente los vectores de entrada X a través del hiperplano, esto se lograra variando los coeficientes del vector de pesos y θ .

2.2.1. Criterio de activación

Consiste en asignar valores de salida a la neurona bajo ciertas restricciones, lo cual permitirá clasificar correctamente los vectores de entrada, para ello la suma ponderada $s = \sum_{j=1}^m w_j x_j - \theta$ será un estimador bajo el siguiente criterio

$$f(s) = \begin{cases} 1 & \sum_{j=1}^m w_j x_j - \theta \geq 0 \\ 0 & \text{eoc} \end{cases} \quad (2.3)$$

Cabe notar que las salidas no siempre están asociados a $\{0, 1\}$ estas pueden ser $\{-1, 1\}$ esto dependerá de el tipo de función de activación a utilizarse.

Básicamente los elementos que forman una neurona son las siguientes

- Los vectores de entrada $X = (x_1, \dots, x_n)$, también conocidos como patrones de entrenamiento, en ocasiones denotado por P .
- Los vectores de pesos $W = (w_1, \dots, w_n)$, que representan los coeficientes de la ecuación $w_1 x_1, \dots, w_n x_n = \theta$ en el espacio R^n .
- La suma ponderada o producto escalar $s = \sum_{j=1}^m w_j x_j - \theta$ que esta asociado a evaluar la ecuación $w_1 x_1, \dots, w_n x_n = \theta$ en el punto X .
- La función de transferencia o activación “ f ”, el cual es un criterio de decisión que permite clasificar correctamente los patrones de entrada.

2.2.2. Funciones de transferencia

En el proceso de aprendizaje la neurona debe cambiar de un estado actual a un nuevo estado constantemente, esto se logra a través de la función de activación “ f ”, su selección depende del objetivo perseguido y la naturaleza del sistema estudiado.

Las funciones de activación más usuales son la escalón, lineal y sigmoidal

Función de transferencia salto o escalón

Matemáticamente esta función se representa mediante el esquema que se puede apreciar en la siguiente figura 2.2:

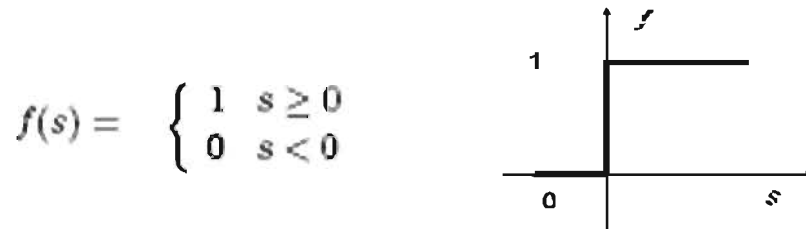


Figura 2.2: Función de transferencia escalón

Fuente: Rich, Eliane y Knight, Kevin

Se observa que la función presenta cambios bruscos en los puntos $(0,0)$, $(0,1)$, precisamente esa región corresponde a un área crítica en la asignación de nuevos valores a el vector de pesos para lograr una correcta clasificación.

Función de transferencia lineal

La función de transferencia lineal se puede observar en la siguiente figura 2.3.



Figura 2.3: Función de transferencia lineal

Fuente: Rich, Eliane y Knight, Kevin

Por lo general esta función es utilizada para reflejar el valor de la suma ponderada sin ninguna modificación, en realidad es la función identidad que es muy útil cuando precisamos conocer el grado de aproximación del vector de entradas respecto del hiperplano.

Función de transferencia sigmoideal

La derivada de la función sigmoideal respecto de s se puede calcular mediante la expresión $\frac{\partial f}{\partial s} = f(1 - f)$, se observa su estructura en el siguiente figura 2.4:

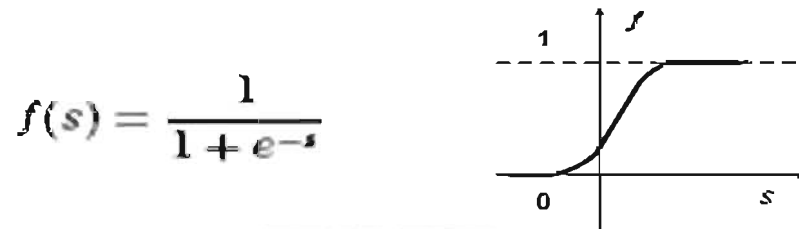


Figura 2.4: Función de transferencia sigmoideal
Fuente: Rich, Eliane y Knight, Kevin

La función de transferencia sigmoideal es una de las funciones más importantes en el estudio de las redes neuronales debido a su comportamiento suave y diferenciable en todo su dominio adecuado para el entrenamiento de redes multicapa, la dificultad con la función de activación escalón surge al no poder evaluar las derivadas en el punto $x = 0$, la función lineal no presenta esta problema pero no realiza ninguna transformación de las entradas por lo tanto no es apropiada para un proceso de entrenamiento.

Debido a esas particularidades la función sigmoide responde adecuadamente a los requerimiento para el entrenamiento de una red multicapa, debido a que es diferenciable en la región crítica $-6 \leq x \leq 6$ donde existe mayor predominancia en el cambio del vector de pesos.

2.3. Topología de las redes neuronales

La topología o arquitectura de las redes neuronales artificiales consiste en la organización y disposición de las neuronas en la red formando capas o agrupaciones de neuronas, en este sentido los parámetros de la red son

- El número de capas
- El número de neuronas por capa
- Tipo de conexión entre neuronas

2.3.1. Redes monocapa

Las redes monocapa están formadas por una sola capa de S neuronas conectados a los R componentes del vector de entradas P , estas pueden tener 1 ó S salidas dependiendo de la naturaleza del problema como se puede apreciar en la siguiente figura 2.5:

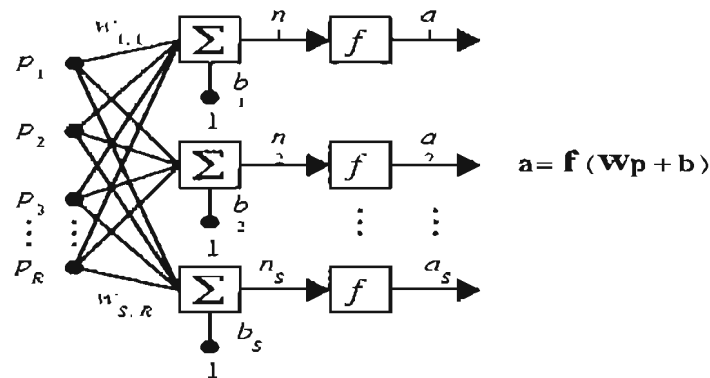


Figura 2.5: Organización de la red monocapa
Fuente: Neural Network Toolbox MATLAB 7.0

La salida de la i -ésima neurona está dada por $a(i) = f(W(i,j)P + b(i))$, donde $W(i,j)$ es el vector de pesos asociado a la i -ésima neurona y j -ésima capa, el vector de entradas denotado por P con R componentes, de forma análoga $b(i)$ es el componente i -ésimo del vector "bias" (θ) que corresponde al umbral, normalmente los elementos de la misma capa tienen la misma función de activación f .

El número de neuronas de una capa no tiene por qué ser igual que el número de componentes del vector de entradas. Generalmente $R \neq S$, entre estos tipos de redes mencionamos a los modelos de Hopfield, Boltzmann Machine.

2.3.2. Redes multicapa

Las redes multicapa son topologías poderosas, debido a que pueden separar regiones convexas en un espacio de dimensión n , la información fluye unidireccionalmente desde las unidades de entrada a las unidades de salida, esta topología fue posible gracias al desarrollo de algoritmos de aprendizaje como el backpropagation.

Una red multicapa se caracteriza por estar claramente identificado las tres regiones o capas

Capa de entrada

Básicamente esta compuesta por los vectores de entrada de la red, esta región no contiene ninguna neurona por lo tanto no se lleva a cabo ningún proceso, algunos autores no lo consideran como una capa.

Capas ocultas

Son aquellas que pueden estar formado por un número considerable o no de neuronas y diferentes conexiones en varios niveles, además para la propagación del error y actualizaciones en los pesos es necesario que las funciones de activación sean continuas.

Capa de salida

Recibe información de la capa oculta para transmitirla al exterior esta puede estar formada por una o varias neuronas de salida dependiendo del tipo de problema a abordar, además las funciones de activación pueden ser de tipo continuas o lineales.

Un esquema general de una red de tres capas se observa en la siguiente figura 2.6

2.3.3. Consideraciones de la red multicapa

- Toda componente del vector de entrada $P = (p_1, \dots, p_R)$ incide en todas las neuronas de la primera capa de oculta.
- El numero de neuronas en la capa j – esima esta dado por S^j .
- Las salidas $f^j = (f_1^j, \dots, f_{s_j}^j)$ de la capa j – esima son el vector de entradas para la capa $j + 1$.
- El vector de pesos en la primera capa oculta es denotado IW de dimenciones $(S^1 \times R)$.

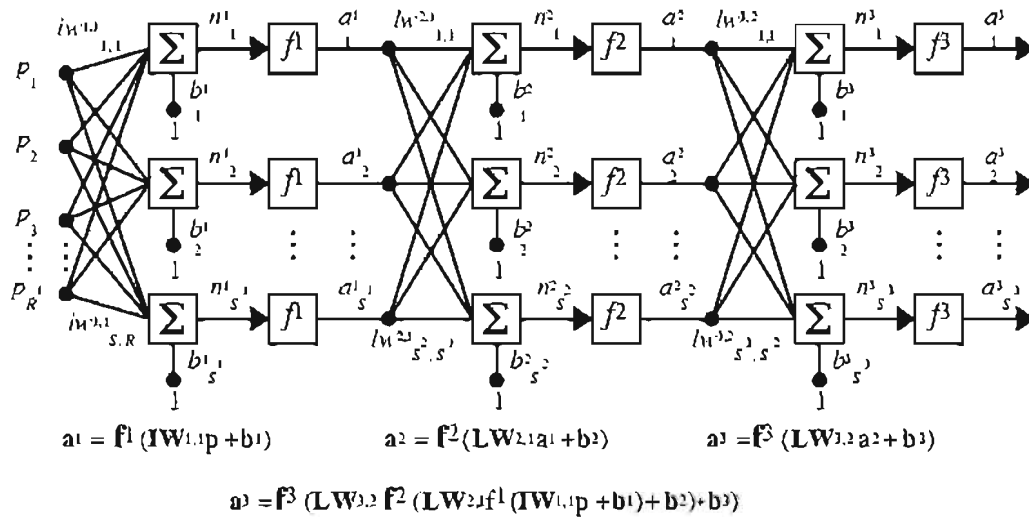


Figura 2.6: Organización de la red multicapa
Fuente: Neural Network Toolbox MATLAB 7.0

- En las demás capas los vectores de pesos se denotan LW de dimensiones $(S^j \times S^{j-1})$ en la capa j – esima respectivamente.
- La notación $lw^{3,2}$ representa el vector de pesos correspondiente a la capa 3 que proviene o depende de la capa 2.
- La suma ponderada o producto escalar en la j – esima capa e i – esima neurona es denotado $n_i^j = a^{j-1} * w_i^j$.
- El umbral asignado a cada neurona, también denominado *bias* esta denotado b_i^j en la j – esima capa e i – esima neurona.
- Las salidas de la red en la ultima capa corresponden a $(a_1^k, \dots, a_{s^k}^k)$.

Desde el punto de vista del flujo de la señal desde una capa a otra podemos establecer una clasificación y distinguir entre dos tipos de redes

- Si las señales y las entradas de las capas intermedias se propagan siempre adelante, la red se denomina estática o red “*feedforward*”, en este caso toda la información se dirige siempre hacia la salida y no hay caminos de vuelta permitidos.
- En cambio, si algunos estados intermedios o salidas de la red son devueltos hacia atrás decimos que la red es *dinámica* o *red recurrente*, aquí las señales son

reutilizadas, es decir las salidas actuales del sistema dependen de las salidas anteriores, estas redes con conexiones hacia adelante y hacia atrás son conocidas como “*feedforward/feedback*”

2.4. Interpretación geométrica

La interpretación geométrica nos provera una visión clara de los procesos de entrenamiento que realiza la red neuronal, a continuación se describen dichos procesos para ello sean los vectores de dimensión n

$$X = (x_1, x_2, \dots, x_n) \quad (2.4)$$

$$W = (w_1, w_2, \dots, w_n) \quad (2.5)$$

x_i son las componentes de los vectores de entrada X , que también son conocidos como patrones de entrenamiento, en ocasiones denotados por P , estos pueden ser de tipo continuo o discreto.

w_i son denominados pesos que representan el nivel de relevancia de la variable x_i , geoméricamente los pesos representan la orientación del hiperplano n -dimensional.

Se denomina conjunto de entrenamiento de la red neuronal a la expresión

$$\lambda = \{(X, T) / X \in R^n, T \in R^m\} \quad (2.6)$$

donde X son los patrones de entrada seleccionados bajo algún criterio estadístico y T es el vector que contiene los valores esperados que pueden ser $\{-1, 1, 0\}$ de acuerdo a la función de activación, m es la cantidad de vectores de entrenamiento.

La red neuronal tiene la propiedad de clasificar un conjunto de puntos dispersos en un espacio de dimensión n a través del hiperplano, de ahí que no es suficiente tener un

control de dirección de este plano sino poder controlar el desplazamiento esto se logra a través del parámetro conocido como **Bias** o **Umbral** “ θ ”, cuya función es controlar el desplazamiento del hiperplano en el espacio R^n .

El producto escalar o suma ponderada representado por la expresión

$$W * X = \sum_{i=1}^n w_i * x_i \quad (2.7)$$

cumple ciertas propiedades tal es el caso del término siguiente

$$W * X = 0 \quad (2.8)$$

que representa un hiperplano que atraviesa el origen de coordenadas en R^n , es decir gráficamente se puede hacer una analogía con el espacio tridimensional entonces quedara representado como se puede apreciar en la siguiente figura 2.7.

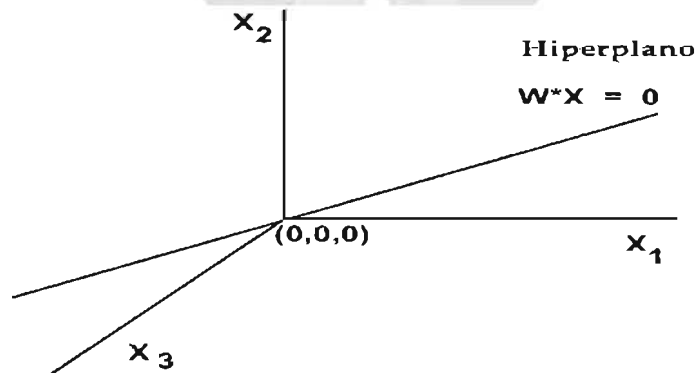


Figura 2.7: Hiperplano en el origen de coordenadas R^3

Fuente: Nils, J. Nilsson

Aplicando el criterio umbral “ θ ”, tal que $\theta > 0$ podemos conseguir el desplazamiento del hiperplano distante del origen, para este propósito se considera la siguiente ecuación

$$W * X = \theta \quad (2.9)$$

Al ser $\theta > 0$ cada componente del producto escalar se dispersa del origen, dando como resultado la siguiente figura 2.8.

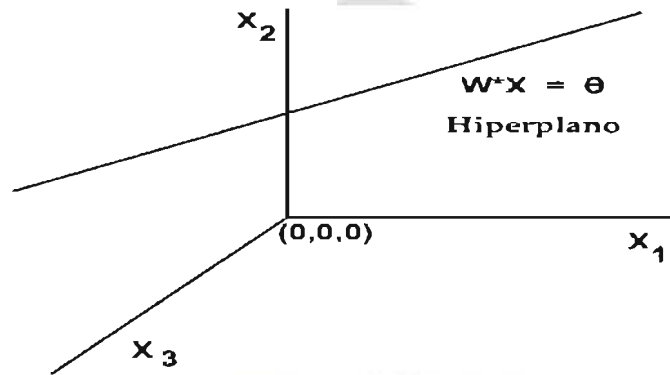


Figura 2.8: Hiperplano disperso respecto del origen de coordenadas R^3
Fuente: Nils, J. Nilsson

Se infiere con lo descrito anteriormente en un clasificador representado por la ecuación $W * X = \theta$ cuyas propiedades en base matemática cumplen las siguientes restricciones

- Existe un conjunto de puntos $W * X - \theta > 0$ que esta claramente definido a un extremo de $W * X - \theta = 0$
- Existe un conjunto de puntos $W * X - \theta < 0$ que esta claramente definido al otro extremo de $W * X - \theta = 0$

Se puede apreciar gráficamente esta propiedad en la siguiente figura 2.9

Observando las propiedades inherentes al grafico el vector de pesos siempre es perpendicular al hiperplano y tiende o converge hacia el vector de entradas X para un correcta clasificación de este patrón.

La idea consiste en ajustar los pesos $W = (w_1, w_2, \dots, w_n)$ de tal forma que el hiperplano pueda diferenciar y clasificar correctamente los conjuntos representados por las ecua-

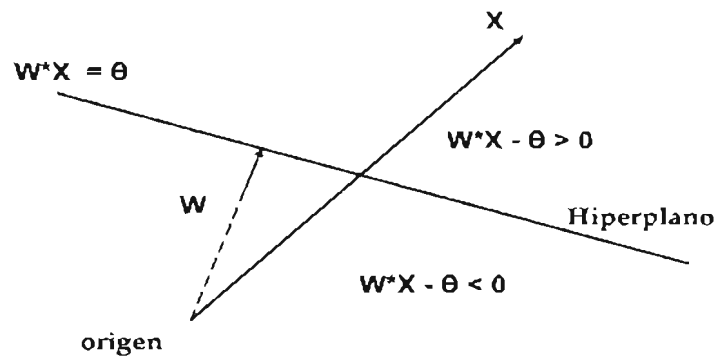


Figura 2.9: Clasificación respecto del Hiperplano

Fuente: Nils, J. Nilsson

ciones $W * X - \theta < 0$, $W * X - \theta > 0$.

Se requiere un criterio de decisión, para asignar valores de salida a los conjuntos $W * X - \theta > 0$ y $W * X - \theta < 0$ e inferir si corresponde al valor deseado, es así que se introduce una función de activación que responderá a los vectores de entrada X con lo cual se da origen a los modelos matemáticos de redes neuronales artificiales.

Realizando una analogía con la neurona biológica $X = (x_1, x_2, \dots, x_n)$ corresponde a las dendritas y $W = (w_1, w_2, \dots, w_n)$ es el valor de inhibición o comunicación entre ellas, el núcleo constituido por la función umbral “ θ ” que la suma ponderada $S = \sum_{i=1}^n W_i * X_i$ debe sobrepasar para poder activarse, finalmente la señal de salida esta formada por la función de activación $f(W * X - \theta)$ que representa al axón.

2.4.1. La dimensión $n+1$

Sin pérdida de generalidad se transforma la expresión $W * X - \theta = 0$ en $W * X = 0$, para ello sea el vector de pesos $W = (w_1, w_2, \dots, w_n)$ de dimensión n , si se agrega una componente más w_{n+1} el nuevo vector de pesos tendrá la siguiente estructura $W = (w_1, w_2, \dots, w_n, w_{n+1})$

La propiedad radica en asignar a la nueva componente el valor del umbral con signo negativo es decir

$$w_{n+1} = -\theta \quad (2.10)$$

Por tanto el vector de pesos resulta

$$W = (w_1, w_2, \dots, w_n, -\theta) \quad (2.11)$$

Bajo el mismo criterio se considera el vector de entradas, relacionando $x_{n+1} = 1$

$$X = (x_1, x_2, \dots, x_n, x_{n+1}) \quad (2.12)$$

$$X = (x_1, x_2, \dots, x_n, 1) \quad (2.13)$$

En estas condiciones se demuestra que $W * X - \theta = 0$, es equivalente a $W * X = 0$, que representa una de las propiedades del estudio de las redes neuronales.

Demostración

Sean los siguientes vectores de dimensión $n + 1$

$$W = (w_1, w_2, \dots, w_n, w_{n+1}) \quad (2.14)$$

$$X = (x_1, x_2, \dots, x_n, x_{n+1}) \quad (2.15)$$

Entonces

$$\begin{aligned}
W * X &= (w_1, w_2, \dots, w_n, w_{n+1})(x_1, x_2, \dots, x_n, x_{n+1}) \\
&= w_1 x_1, w_2 x_2, \dots, w_{n+1} x_{n+1} \\
&= \sum_{i=1}^n w_i * x_i + w_{n+1} x_{n+1} \\
&= \sum_{i=1}^n w_i * x_i - \theta \\
&= W * X - \theta
\end{aligned} \tag{2.16}$$

En lo sucesivo se considerara que todos los vectores son construidos de la misma forma, bajo esta nueva notación el Hiperplano n -dimensional esta dado por la ecuación $W * X = 0$ que permite la clasificación de los valores $W * X > 0$ en un extremo del Hiperplano y su complemento $W * X < 0$ en el otro.

2.5. Algoritmos de aprendizaje

Consiste en los cambios de comportamiento inducido por la interacción con el entorno; Biológicamente la información memorizada esta relacionada con los valores sinápticos de las conexiones entre las neuronas, en el caso de las neuronas artificiales el conocimiento se encuentra almacenado en los pesos, el aprendizaje consiste en la variación de los pesos.

El aprendizaje depende de la arquitectura o topología de la red y del tipo de conexión auto recurrente o con conexiones laterales , cada modelo dispone de sus propias técnicas de aprendizaje.

2.5.1. Método del gradiente descendente

Para todo los vectores del conjunto de entrenamiento denotado por la expresión

$$\lambda = \{(X, T) / X \in \mathbb{R}^n, T \in \mathbb{R}^m\} \tag{2.17}$$

le corresponden una o varias salidas de la red, denotados f_i , análogamente el vector T contiene los valores deseados d_i .

El criterio que nos permite medir cuan próximo es el valor obtenido por la red neuronal respecto del valor deseado será la función del error cuadrático promedio mínimo denotado

$$\varepsilon = \frac{1}{m} \sum_{X \in \lambda} (d_i - f_i)^2 \quad (2.18)$$

Obsérvese que la sumatoria se extiende a todos los vectores de entrenamiento, además el error esta asociado directamente a los vectores de pesos " W ", la idea consiste en que el error sea cero o conberga a este valor, esta aproximación en la mayoría de los casos es buena.

El error " ε " depende de todos los vectores de entrenamiento de ahí que se debe aplicar un proceso incremental de entrenamiento siguiendo alguna determinada secuencia, es decir seleccionar un único vector de entrada a la red hacer los ajustes correspondientes a los pesos para obtener la respuesta adecuada y volver a repetir el proceso.

Bajo este criterio el error para un único vector de entradas se define

$$\varepsilon = (d_i - f_i)^2 \quad (2.19)$$

Aplicando la función gradiente al error, que consiste en extraer derivadas parciales respecto de los pesos se tiene

$$\frac{\partial \varepsilon}{\partial W} = \left[\frac{\partial \varepsilon}{\partial w_1} \cdots \frac{\partial \varepsilon}{\partial w_i} \cdots \frac{\partial \varepsilon}{\partial w_{n+1}} \right] \quad (2.20)$$

El error “ ε ” depende de los pesos W a través del producto escalar $S = W * X$, entonces aplicando la regla de derivación de la cadena obtenemos

$$\frac{\partial \varepsilon}{\partial W} = \frac{\partial \varepsilon}{\partial S} \frac{\partial S}{\partial W} \quad (2.21)$$

Además la derivada parcial de $S = W * X$ respecto de W esta dada por $\frac{\partial S}{\partial W} = X$, con lo que la expresión 2.21 se puede escribir

$$\frac{\partial \varepsilon}{\partial W} = \frac{\partial \varepsilon}{\partial S} X \quad (2.22)$$

Básicamente se busca la variabilidad del error “ ε ” respecto de los pesos, para ello solamente resta deducir la variabilidad del error respecto del producto escalar o suma ponderada $S = W * X$, para ello consideremos la expresión $\varepsilon = (d_i - f_i)^2$ aplicando la derivada respecto de S se obtiene

$$\frac{\partial \varepsilon}{\partial S} = -2(d - f) \frac{\partial f}{\partial S} \quad (2.23)$$

Finalmente la ecuación 2.21 que representa los cambios del error respecto de los pesos puede escribirse

$$\frac{\partial \varepsilon}{\partial W} = -2(d - f) \frac{\partial f}{\partial S} X \quad (2.24)$$

Los cambios en S son casi imperceptibles, pero cuando “ f ” cambia lo hace de manera abrupta de 0 a 1, o viceversa, bajo estas condiciones se requiere alguna función “ f ” con estas características, existen dos formas de solventar esta dificultad

- Asumir que $f = S$
- Asignar a f alguna función no lineal que sea diferenciable

2.5.2. Procedimiento Widrow-Hoff

Consiste en asignar “ $f = S$ ” del tal forma que el error cuadrático puede ser representado por $\varepsilon = (d - S)^2$, además “ $f = S$ ” representa la función identidad, es decir la salida corresponderá a algún valor dado por la suma ponderada $S = W * X$ cuyos resultados oscilan en los números reales.

Para poder realizar el proceso de clasificación se requieren valores deseados d_i que permanezcan en alguno de los estados posibles $\{-1, 0, 1\}$ los cuales permitirán realizar un correcto proceso de clasificación.

La función de activación definida por la siguiente expresión es esencial en el proceso de clasificación

$$f(s) = \begin{cases} 1 & \sum_{i=1}^{n+1} w_i x_i \geq 0 \\ 0 & \text{eoc} \end{cases} \quad (2.25)$$

Si “ $f = S$ ” entonces $\frac{\partial f}{\partial S} = 1$ con lo que el gradiente del error cuadrático puede ser escrito de la siguiente manera

$$\frac{\partial \varepsilon}{\partial W} = -2(d - f)X \quad (2.26)$$

La anterior expresión representa los cambios en el error, debido a variaciones en el vector de pesos, en términos operacionales esta dada por la expresión $-2(d - f)X$, donde la constante numérica -2 , es un factor que propaga la expresión $(d - f)X$ duplicando esta por su valor original.

Sea η un elemento de los números reales en particular $\eta = -2$, cuando η tiende a crecer la expresión $(d - f)X$ se propaga geoméricamente $\eta - veces$ y recíprocamente cuando η tiende a decrecer esta se contrae, esto incide en la rapidez de cambios en el error cuadrático.

Es así que η es denominado factor o velocidad de aprendizaje, en consecuencia la ecuación 2.26 puede expresarse:

$$\frac{\partial \varepsilon}{\partial w} = \eta(d - f)X \quad (2.27)$$

Estudios realizados demostraron que, para que una red alcance la convergencia y no pierda estabilidad el factor de aprendizaje óptimo debería oscilar en el intervalo

$$0 < \eta < 2$$

De manera simultanea Rosemblatt trabajo con el modelo del perceptron mientras Bernard Widrow y su estudiante Marcian Hoff desarrollaron el modelo Adaline y su regla de aprendizaje denominado "*Least Mean Square*" (*LMS*).

El algoritmo "*LMS*" es mas potente que la regla de aprendizaje del perceptron porque minimiza el error medio cuadrático para todos los vectores del conjunto de entrenamiento, este es el gran aporte de esta red.

Como se dedujo anteriormente la variación del error esta ligada a la variabilidad de los pesos a través del producto escalar $S = W * X$ de ahí es natural asumir que el aprendizaje se realice través de la siguiente expresión:

$$W = W + n(d - f)X \quad (2.28)$$

Además se debe considerar las siguientes propiedades que son necesarias para comprender bajo que condiciones se realizan la actualización de los pesos

$$W = \begin{cases} W & \text{si } (d - f) = 0 \\ \text{Expande} & \text{si } (d - f) > 0 \\ \text{Contrae} & \text{si } (d - f) < 0 \end{cases} \quad (2.29)$$

2.5.3. Procedimiento delta generalizado

El problema de la no derivabilidad de la función umbral debido a los cambios abruptos, condujo a la búsqueda de funciones diferenciables cuya salidas oscilen entre los conjuntos $\{1, -1\}$ y $\{0, 1\}$ es así que se consolidaron las funciones: Gaussiana, Tangente hiperbólica, Sinusoidal y Logística o logsigmoidea.

Función gaussiana

$$f(s) = \exp(-s^2) \quad (2.30)$$

Función Tangente hiperbólica

$$\tanh(s) = \frac{\exp(s) - \exp(-s)}{\exp(s) + \exp(-s)} \quad (2.31)$$

Función sinusoidal

$$f(s) = \text{sen}(s) \quad (2.32)$$

Para el presente estudio utilizaremos la función logística o logsigmoidea, que esta definida

$$f(s) = \frac{1}{1 + e^{-s}} \quad (2.33)$$

cuya derivada puede calcularse en términos de si misma y , corresponde a

$$\frac{\partial f}{\partial s} = f(1 - f) \quad (2.34)$$

Demostración

$$\begin{aligned} \frac{\partial f}{\partial s} &= \frac{\partial\left(\frac{1}{1+e^{-s}}\right)}{\partial s} \\ &= \frac{(1+e^{-s})\frac{\partial 1}{\partial s} - \frac{\partial(1+e^{-s})}{\partial s}}{(1+e^{-s})^2} \\ &= \frac{0 - e^{-s}(-1)}{(1+e^{-s})^2} \\ &= \frac{e^{-s}}{(1+e^{-s})^2} \\ &= f\left(1 - \frac{1}{1+e^{-s}}\right) \\ &= f(1 - f) \end{aligned} \quad (2.35)$$

Recordemos que el gradiente del error cuadrático esta dado por

$$\frac{\partial \varepsilon}{\partial W} = \eta(d - f) \frac{\partial f}{\partial S} X \quad (2.36)$$

Sustituyendo la derivada de la función sigmoide en el gradiente se obtiene

$$\frac{\partial \varepsilon}{\partial W} = \eta(d - f)f(1 - f)X \quad (2.37)$$

De manera que se puede deducir la regla de cambios de los pesos, esta regla es conocida como el procedimiento delta generalizado.

$$W = W + \eta(d - f)f(1 - f)X \quad (2.38)$$

2.6. Metodo Backpropagation o retropropagación

La arquitectura multicapa ofrecía grandes ventajas al poder representar una complejidad mayor de los patrones de entrada, la dificultad surgió al no contar con algoritmos que permitieran entrenar redes de esta naturaleza es así que durante mucho tiempo el campo de las redes neuronales quedó detenido.

A continuación se describe el proceso de retropropagación, dado un conjunto inicial de entrenamiento esta se propaga por toda la red en todas sus capas y neuronas hasta hallar la salida, una vez obtenido el resultado se compara con el valor esperado hallando generalmente un margen de error, este debe ser corregido a través de la red modificando los pesos.

Es así que surge el método “*backpropagation*” ideado a principios de los 70 por Herbos y redescubierto a principios de los 80 por Parker Rumelhart independientemente.

La idea consiste en propagar el error hacia atrás, es decir de la capa de salida hacia la capa de entrada pasando por las capas ocultas en el proceso ajustando los pesos para minimizar el error.

Consideremos como base inicial el error cuadrático para desarrollar el método de la “*Retropropagación*”, entonces sea el término

$$\varepsilon = (d_i - f_i)^2 \quad (2.39)$$

Que correspondiente a la salida i – *esima* en una arquitectura con múltiples salidas.

El gradiente del error en la unidad sigmoideal i en la capa j – *esima* esta dado por

$$\frac{\partial \varepsilon}{\partial W_i^j} = \left[\frac{\partial \varepsilon}{\partial w_{1i}^j} \cdots \frac{\partial \varepsilon}{\partial w_{ri}^j} \cdots \frac{\partial \varepsilon}{\partial w_{(m_{j-1}+1)i}^j} \right] \quad (2.40)$$

Respecto a las componentes r – esimas oscilan en el intervalo $(1 \leq r \leq m_{j-1} + 1)$ y corresponden al vector de pesos en la unidad sigmoideal i en la capa j .

Puede deducirse que el error depende de todas las capas y neuronas por tanto " ε " depende de W_i^j , aplicando la regla de derivación de la cadena se tiene

$$\frac{\partial \varepsilon}{\partial W_i^j} = \frac{\partial \varepsilon}{\partial S_i^j} \frac{\partial S_i^j}{\partial W_i^j} \quad (2.41)$$

Debido a que

$$S_i^j = X^{j-1} W_i^j \quad (2.42)$$

Y además la derivada parcial del producto escalar respecto de los pesos es

$$\frac{\partial S_i^j}{\partial W_i^j} = X^{j-1} \quad (2.43)$$

Sustituyendo en la expresión 2.41 se obtiene

$$\frac{\partial \varepsilon}{\partial W_i^j} = \frac{\partial \varepsilon}{\partial S_i^j} X^{j-1} \quad (2.44)$$

Al depender “ ε ” de W_i^j a través de S_i^j , la derivada parcial esta dada por la siguiente expresión

$$\begin{aligned}\frac{\partial \varepsilon}{\partial S_i^j} &= \frac{\partial (d-f)^2}{\partial S_i^j} \\ &= -2(d-f) \frac{\partial f}{\partial S_i^j}\end{aligned}\quad (2.45)$$

Sustituyendo en 2.46 la expresión obtenida en 2.44 encontramos finalmente

$$\frac{\partial \varepsilon}{\partial W_i^j} = -2(d-f) \frac{\partial f}{\partial S_i^j} X^{j-1} \quad (2.46)$$

Efectuando operaciones resulta

$$\begin{aligned}\frac{\partial \varepsilon}{\partial W_i^j} &= -2(d-f) \frac{\partial f}{\partial S_i^j} X^{j-1} \\ \frac{\partial \varepsilon}{\partial W_i^j} \frac{1}{-2X^{j-1}} &= (d-f) \frac{\partial f}{\partial S_i^j} \\ \frac{\partial \varepsilon}{-2\partial W_i^j X^{j-1}} &= (d-f) \frac{\partial f}{\partial S_i^j} \\ \frac{\partial \varepsilon}{-2\partial S_i^j} &= (d-f) \frac{\partial f}{\partial S_i^j}\end{aligned}\quad (2.47)$$

Finalmente

$$\frac{\partial \varepsilon}{-2\partial S_i^j} = (d-f) \frac{\partial f}{\partial S_i^j} \quad (2.48)$$

La expresión 2.48 desempeña un papel muy importante en la propagación del error, para simplificarla nos referiremos a ella por “ δ_i^j ” que representa la sensibilidad del error cuadrático de la salida de la red respecto a los cambios en la entrada de la correspondiente

función sigmoide, de esta forma el gradiente del error se puede escribir como

$$\frac{\partial \varepsilon}{\partial W_i^j} = -2\delta_i^j X^{j-1} \quad (2.49)$$

Puesto que debemos modificar los pesos en sentido negativo del gradiente e introduciendo el factor de aprendizaje η_i^j , la actualización de los pesos responde a la siguiente expresión

$$W_i^j = W_i^j + \eta_i^j \delta_i^j X^{(j-1)} \quad (2.50)$$

Normalmente el factor de aprendizaje es el mismo para todos los vectores de pesos aunque es posible trabajar con diferentes factores de aprendizaje en cada capa y neurona.

2.6.1. Análisis de “ δ_i^j ” en la última capa

La sensibilidad del error cuadrático “ δ_i^j ” se propaga a través de toda la red, iniciando en la capa de salida y realizando un proceso de retropropagación a través de las capas ocultas hacia la capa de entrada, dicha sensibilidad operacionalmente se representa en la capa de salida por la siguiente expresión

$$\delta^k = (d - f) \frac{\partial f}{\partial s^k} \quad (2.51)$$

Además la derivada parcial de la función sigmoide respecto del producto escalar o suma ponderada $S^k = X^{(k-1)}W^k$ esta dada por $f(1 - f)$ reemplazando este valor en la expresión 2.51 resulta

$$\delta^k = (d - f)f(1 - f) \quad (2.52)$$

La ecuación anterior representa la sensibilidad del error cuadrático en la capa de salida, este resultado es importante ya que permite ajustar los pesos en función al error en la última capa, este procedimiento se denomina *retropropagación*, el cual puede ser escrita

$$W^k = W^k + \eta^k (d - f) f(1 - f) X^{(k-1)} \quad (2.53)$$

2.6.2. Análisis de “ δ_i^j ” en las capas intermedias

La idea básicamente consiste en utilizar la sensibilidad del error cuadrático calculada en la última capa para cambiar el vector de pesos en las capas intermedias, para ello recordemos que

$$\delta_i^j = (d - f) \frac{\partial f}{\partial s_i^j} \quad (2.54)$$

Siendo que el error se propaga hacia atrás se crea una dependencia funcional bajo el siguiente criterio f_i^j , depende de $S_q^{(j+1)}$ y $S_q^{(j+1)}$ depende de S_i^j , es decir

$$f_i^j(s_q^{(j+1)}(s_i^j)) \quad (2.55)$$

El grafico 2.10 ilustra el proceso de dependencia funcional

Aplicando la regla de derivación de la cadena se deduce

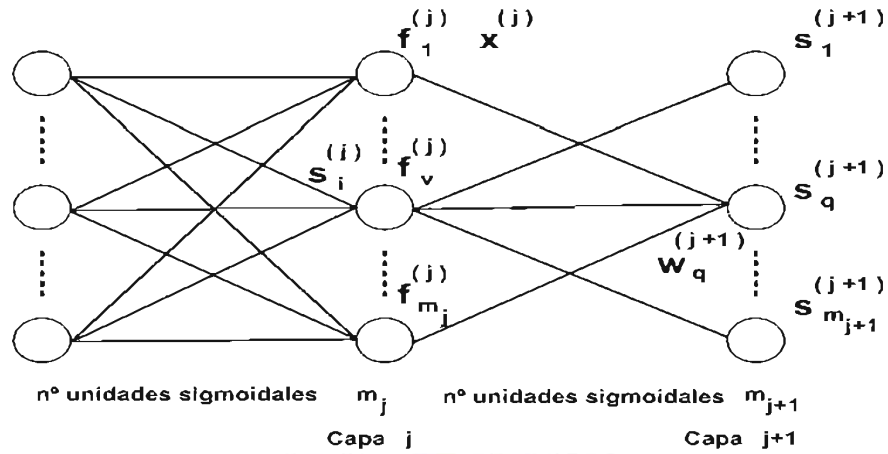


Figura 2.10: Dependencia funcional de la arquitectura de retropropagación
Fuente: Nils, J. Nilsson

$$\begin{aligned}
 \tilde{\delta}_i^j &= (d - f) \left[\frac{\partial f}{\partial s_1^{(j+1)}} \frac{\partial s_1^{(j+1)}}{\partial s_i^j} + \dots + \frac{\partial f}{\partial s_q^{(j+1)}} \frac{\partial s_q^{(j+1)}}{\partial s_i^j} + \dots + \frac{\partial f}{\partial s_{m_{j+1}}^{(j+1)}} \frac{\partial s_{m_{j+1}}^{(j+1)}}{\partial s_i^j} \right] \\
 &= \sum_{q=1}^{m_{j+1}} (d - f) \frac{\partial f}{\partial s_q^{(j+1)}} \frac{\partial s_q^{(j+1)}}{\partial s_i^j} \\
 &= \sum_{q=1}^{m_{j+1}} \delta_q^{(j+1)} \frac{\partial s_q^{(j+1)}}{\partial s_i^j}
 \end{aligned} \tag{2.56}$$

Por lo tanto

$$\delta_i^j = \sum_{q=1}^{m_{j+1}} \delta_q^{(j+1)} \frac{\partial s_q^{(j+1)}}{\partial s_i^j} \tag{2.57}$$

Bien solo queda calcular las derivadas parciales de la expresión $\frac{\partial s_q^{(j+1)}}{\partial s_i^j}$ para ello hay que tener en cuenta

$$s_q^{(j+1)} = \sum_{r=1}^{m_j+1} f_r^j w_{rq}^{(j+1)} \tag{2.58}$$

$$\begin{aligned}
S_q^{(j+1)} &= X^j W_q^{(j+1)} \\
&= (f_1^j, \dots, f_i^j, \dots, f_{m_j}^j, 1)(w_{1q}^{(j+1)}, \dots, w_{m_j q}^{(j+1)}, -\theta) \\
&= \sum_{r=1}^{m_j+1} f_r^j w_{rq}^{(j+1)}
\end{aligned} \tag{2.59}$$

Finalmente

$$S_q^{(j+1)} = \sum_{r=1}^{m_j+1} f_r^j w_{rq}^{(j+1)} \tag{2.60}$$

Ahora interesa saber como cambia la suma ponderada $S_q^{(j+1)}$ en la capa $j + 1$ en la unidad sigmoideal q respecto de la suma ponderada S_i^j en la capa j de la unidad sigmoideal i para este propósito derivamos la igualdad anterior

$$\frac{\partial S_q^{(j+1)}}{\partial S_i^j} = \frac{\partial(\sum_{r=1}^{m_j+1} f_r^j w_{rq}^{(j+1)})}{\partial S_i^j} \tag{2.61}$$

y si tomamos en cuenta que los pesos no dependen de S entonces

$$\frac{\partial S_q^{(j+1)}}{\partial S_i^j} = \sum_{r=1}^{m_j+1} w_{rq}^{(j+1)} \frac{\partial f_r^j}{\partial S_i^j} \tag{2.62}$$

además en el proceso de conexiones hacia adelante “feedforward” se tiene la dependencia funcional

$$f_r^j(s_i^j) \tag{2.63}$$

en la capa j – esima, se puede observar que la entrada S_i^j solamente incide en f_r^j siempre que $r = i$ bajo estas condiciones.

$$\frac{\partial f_r^j}{\partial S_i^j} = \begin{cases} 0 & r \neq i \\ f_i^j(1 - f_i^j) & r = i \end{cases} \quad (2.64)$$

por lo tanto la ecuación 2.62 se puede escribir

$$\begin{aligned} \frac{\partial s_q^{(j+1)}}{\partial s_i^j} &= \sum_{\tau=1}^{m_j+1} w_{\tau q}^{(j+1)} \frac{\partial f_{\tau}^j}{\partial s_i^j} \\ &= w_{1q}^{(j+1)} \frac{\partial f_1^j}{\partial s_i^j} + \dots + w_{iq}^{(j+1)} \frac{\partial f_i^j}{\partial s_i^j} + \dots + w_{(m_j+1)q}^{(j+1)} \frac{\partial f_{m_j+1}^j}{\partial s_i^j} \\ &= 0 + \dots + w_{iq}^{(j+1)} f_i^j(1 - f_i^j) + \dots + 0 \end{aligned} \quad (2.65)$$

De la ecuación anterior se deduce

$$\frac{\partial s_q^{(j+1)}}{\partial S_i^j} = w_{iq}^{(j+1)} f_i^j(1 - f_i^j) \quad (2.66)$$

Finalmente la sensibilidad del error cuadrático queda expresado en términos recursivos dada por la ecuación

$$\begin{aligned} \delta_i^j &= \sum_{q=1}^{m_j+1} \delta_q^{(j+1)} \frac{\partial s_q^{(j+1)}}{\partial s_i^j} \\ &= \sum_{q=1}^{m_j+1} \delta_q^{(j+1)} w_{iq}^{(j+1)} f_i^j(1 - f_i^j) \\ &= f_i^j(1 - f_i^j) \sum_{q=1}^{m_j+1} \delta_q^{(j+1)} w_{iq}^{(j+1)} \end{aligned} \quad (2.67)$$

Como se puede apreciar la ecuación anterior es recursiva, resulta interesante resaltar que la expresión no depende de los cambios en la función de activación, solo afecta explícitamente al cálculo de “ δ^k ” en la última capa.

Una vez calculado “ $\delta_q^{(j+1)}$ ” en la capa $j+1$ podemos utilizar este resultado para calcular “ δ_i^j ” en la capa j .

La base de la recurrencia es “ δ^k ” que puede ser calculado por medio de la expresión

$$\delta^k = (d - f)f(1 - f) \quad (2.68)$$

Finalmente:

$$\delta_i^j = \begin{cases} f_i^j(1 - f_i^j) \sum_{q=1}^{m_{j+1}} \delta_q^{(j+1)} w_{iq}^{(j+1)} & \text{si } j \neq k \\ (d - f_i^j)f_i^j(1 - f_i^j) & \text{si } j = k \end{cases} \quad (2.69)$$

Usando la expresión que hemos obtenido el valor de la sensibilidad del error cuadrático “ δ_i^j ”, la regla genérica para el cambio de los pesos puede ser escrita

$$W_i^{(j)} = W_i^{(j)} + n_i^{(j)} \delta_i^{(j)} X^{(j-1)} \quad (2.70)$$

Básicamente el principio de la *retropropagación* consiste en propagar “ δ_i^j ” por todos los pesos de la red, desde la última capa hacia la primera y evaluando el error cuadrático medio $\varepsilon = \frac{\sum_{i=1}^N (d_i^k - f_i^k)^2}{N}$ el cual si converge a cero, la red neuronal habrá clasificado correctamente los patrones de entrada.

Capítulo 3

Series Temporales

3.1. Introducción

Los modelos dinámicos de series temporales permiten estudiar las características de la estructura probabilística de los datos, generalmente los datos reales no tienen comportamiento estacionario es decir no varían alrededor de una varianza constante y una media cero, los modelos “ $ARIMA(p, d, q)$ ” surgen precisamente para el estudio de dichos comportamientos, se originaron a partir de los años (1970) merced a la obra de los estadísticos Box y Jenkis. Un modelo ARIMA se deriva en los modelos autorregresivos $AR(p)$ y medias móviles $MA(q)$, incorporando una escala de integración que permite transformar a un modelo estacionario.

El objetivo principal de esta metodología se refieren a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental.

3.2. Proceso estocástico

Un proceso estocástico es una sucesión de variables aleatorias y_t ordenadas, pudiendo tomar “ t ” cualquier valor entre $-\infty$ y $+\infty$, el subíndice “ t ” en el contexto de análisis de series temporales representa el paso del tiempo.

$$y_{t_1}, y_{t_2}, \dots, y_{t_n} \quad (3.1)$$

Cada variable aleatoria y_{t_i} tendrá su propia función de distribución probabilística $F(y_{t_i})$ con sus correspondientes momentos, de manera análoga si fijamos dos valores en el tiempo, se obtendría una variable bidimensional con una función de distribución bivariante y su función conjunta será $F(y_{t_i}, y_{t_j})$.

En general para un conjunto finito de valores en el tiempo t_1, \dots, t_n se obtiene una función de distribución conjunta

$$F(y_{t_1}, y_{t_2}, \dots, y_{t_n}) \quad (3.2)$$

Se dice que un proceso estocástico está perfectamente caracterizado cuando se pueden determinar las funciones de distribución conjunta para cada conjunto finito de variables del proceso.

Se pueden calcular momentos de diverso orden, aunque son los momentos de primer y segundo orden los más utilizados la media, varianza, covarianza y los coeficientes de correlación

- $\mu_t = E[y_t]$ el subíndice t señala que la media en general será distinta para cada periodo de tiempo
- $\gamma_{t,s} = Cov[y_t, y_s] = E[(y_t - \mu_t)(y_s - \mu_s)]$ en distintos momentos del tiempo
- Cuando $t = s$ se tiene definida la varianza $\gamma_{t,t} = Var[y_t] = E[(y_t - \mu_t)^2]$
- $\rho_{t,s} = \frac{Cov(y_t, y_s)}{\sqrt{Var(y_t)}\sqrt{Var(y_s)}}$ coeficiente de autocorrelación

Las funciones de distribución de probabilidades podrían no estar completamente caracterizadas, los momentos podrían no coincidir incluso no existir para algunas de las variables

aleatorias, lo mismo puede ocurrir con las distribuciones conjuntas o marginales, sin embargo de todos los tipos de procesos estocásticos posibles nos interesa especialmente dos de ellos a los que la estadística a dado nombres precisos

- Ruido blanco
- Proceso estocástico estacionario

3.2.1. Ruido blanco

Es una sucesión de variables aleatorias $\varepsilon_{t_1}, \varepsilon_{t_2}, \dots, \varepsilon_{t_n}$ con esperanza cero, varianza constante e independientes para distintos valores de t , covarianza nula.

$$\begin{aligned} E[\varepsilon_t] &= 0 \quad \forall t \\ \text{Var}[\varepsilon_t] &= E[\varepsilon_t^2] = \sigma^2 \quad \forall t \\ \text{Cov}[\varepsilon_t, \varepsilon_{t'}] &= E[\varepsilon_t \varepsilon_{t'}] = 0 \quad \forall t \neq t' \end{aligned} \quad (3.3)$$

En otros términos el ruido aleatorio ε_t tiene una distribución normal con media cero y varianza constante

$$\varepsilon_t \sim N(0, \sigma^2) \quad (3.4)$$

3.2.2. Proceso estocástico estacionario

Decimos que un proceso estocástico es estacionario si las funciones de distribución conjuntas son invariantes con respecto a un desplazamiento en el tiempo.

Es decir sean los periodos sucesivos de tiempo

$$t, t+1, t+2, \dots, t+k \quad (3.5)$$

Para todo t, k, m no nulos positivos, si el proceso es estacionario se deberá verificar la siguiente igualdad

$$F(y_t, y_{t+1}, \dots, y_{t+k}) = F(y_{t+m}, y_{t+1+m}, \dots, y_{t+k+m}) \quad (3.6)$$

Esta definición de estacionariedad se conoce como *estacionariedad en sentido estricto o fuerte*.

El concepto de estacionariedad será más limitado (amplio o débil), si se verifican las siguientes condiciones

- La esperanza matemática o media teórica permanece constante a lo largo del tiempo.

$$E[y_t] = \mu \quad \forall t$$

- Las varianzas tampoco dependen del tiempo y son finitas.

$$Var[y_t] = Var[y_{t+m}] = E[(y_t - \mu)^2] = \sigma^2 \neq \infty \quad \forall t$$

- La autocovarianza entre dos periodos distintos de tiempo únicamente vienen afectada por el lapso de tiempo transcurrido entre esos dos periodos

$$Cov[y_t, y_{t+k}] = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k \quad \forall t$$

De la última condición depende, si un fenómeno es estacionario, sus variables pueden estar linealmente relacionadas entre si, pero solo dependen de la distancia temporal k transcurrida entre ellas.

Si un proceso es estacionario en sentido estricto también será estacionario en sentido amplio o débil, pero no ocurrirá necesariamente el caso recíproco, pues el proceso puede no ser estacionario para momentos superiores al segundo orden.

3.3. Modelos autorregresivos AR

Los modelos autorregresivos son modelos matemáticos que explican el comportamiento de las observaciones en términos de sí mismas correspondientes a periodos pasados, es decir toda variable aleatoria y_t puede expresarse como una combinación lineal de sus valores pasados más un término de error.

Los modelos autorregresivos se denotan $AR(p)$, el parámetro “ p ” es conocido como factor de retardo expresa el número de observaciones retrasadas de la serie temporal.

3.3.1. Modelo AR(1)

Un modelo $AR(1)$ viene definido por la siguiente ecuación

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \quad (3.7)$$

Se define la propiedad $y_{t-n} = L^n y_t$, en el cual toda observación en un periodo n anterior a t puede ser expresada en términos de una constante L^n que al multiplicar a la observación y_t esta puede ser equivalente a la observación y_{t-n} , siendo así y efectuando operaciones a la ecuación 3.7 resulta

$$(1 - \phi_1 L)y_t = \varepsilon_t \quad (3.8)$$

Es importante denotar la relación que existe entre el ruido blanco ε_t y la variable y_t , ε_t influye en todas las observaciones y_t correspondientes al mismo tiempo o a periodos posteriores, pero nunca ejerce influencia sobre periodos anteriores.

Bajo los criterios mencionados se deduce la covarianza que mide el grado de relación entre las variables ε_t , $y_{t-\tau}$

$$\text{Cov}(\varepsilon_t, y_{t-r}) = E[(\varepsilon_t - 0)(y_{t-r} - \mu)] = 0 \quad \forall r > 0 \quad (3.9)$$

Para que el proceso definido en 3.7 sea estacionario, las raíces de la ecuación característica

$$1 - \phi_1 L = 0 \quad (3.10)$$

deben caer fuera del círculo de la unidad es decir $|\phi_1| < 1$.

Para analizar el comportamiento de la ecuación 3.7 se asume que el proceso se ha iniciado en un periodo $-N$, realizando sustituciones sucesivas resulta

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \dots \\ &= \phi_1^N y_{t-N} + \sum_{k=0}^{N-1} \phi_1^k \varepsilon_{t-k} \end{aligned} \quad (3.11)$$

Evaluando la esperanza se obtiene

$$E(y_t) = \phi_1^N y_{t-N} \quad (3.12)$$

Si $|\phi_1| < 1$ la esperanza decrecerá, todo lo contrario ocurrirá cuando $|\phi_1| > 1$, si $|\phi_1| = 1$ la esperanza se mantiene constante $E(y_t) = y_{t-N}$, y si fuere $|\phi_1| = -1$ se daría una alternancia de signos en el valor de la esperanza.

Un proceso iniciado en $-N$ tendrá la siguiente varianza

$$\begin{aligned}
 \gamma_0 &= E[y_t - E(y_t)]^2 \\
 &= E[y_t - \phi_1^N y_{t-N}]^2 \\
 &= E\left[\sum_{k=0}^{N-1} \phi_1^k \varepsilon_{t-k}\right]^2 \\
 &= E[\varepsilon_{t-k}^2] \sum_{k=0}^{N-1} \phi_1^{2k} \\
 &= \sigma_\varepsilon^2 \sum_{k=0}^{N-1} \phi_1^{2k} \\
 &= \sigma_\varepsilon^2 \frac{1 - \phi_1^{2N}}{1 - \phi_1^2}
 \end{aligned} \tag{3.13}$$

Si el proceso es estacionario es decir $|\phi_1| < 1$ y se ha iniciado en $-\infty$ se verifica la siguiente igualdad

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \tag{3.14}$$

La ecuación anterior corresponde a la varianza para un proceso autorregresivo de orden uno, cuando el proceso arranca en $-\infty$ y cumple la condición de estacionariedad.

Análogamente se obtiene la covarianza respecto a un periodo r multiplicando y_{t-r} al modelo $y_t = \phi_1 y_{t-1} + \varepsilon_t$, y aplicando esperanzas resulta

$$\begin{aligned}
 E[y_t y_{t-r}] &= E[(\phi_1 y_{t-1} + \varepsilon_t) y_{t-r}] \\
 &= \phi_1 \text{Cov}[y_{t-1}, y_{t-r}] + E[\varepsilon_t y_{t-r}] \\
 &= \phi_1 \gamma_{r-1} + 0 \\
 &= \gamma_r
 \end{aligned} \tag{3.15}$$

La expresión deducida resulta de aplicar la covarianza entre las observaciones y_{t-1}, y_{t-r} cuyo tiempo transcurrido esta dado por $|(t-1) - (t-r)| = r-1$, además el ruido aleatorio no puede afectar a una variable en un tiempo menor a el es decir $E[\varepsilon_t y_{t-r}] = 0$.

Dividiendo la expresión 3.15 entre la varianza γ_0 obtenemos los coeficientes de auto correlación para intervalos de tiempo r

$$\begin{aligned} \frac{\gamma_r}{\gamma_0} &= \phi_1 \frac{\gamma_{r-1}}{\gamma_0} \\ \rho_r &= \phi_1 \rho_{r-1} \end{aligned} \quad (3.16)$$

Realizando sustituciones sucesivas se obtiene

$$\rho_r = \phi_1^r \rho_0 \quad (3.17)$$

En las autocovarianzas la condición inicial viene dada por el valor de la varianza , en cambio en las autocorrelaciones como son medidas relativas la condición inicial es siempre $\rho_0 = 1$, así la ecuación 3.17 resulta

$$\rho_r = \phi_1^r, \quad \forall r > 0 \quad (3.18)$$

3.3.2. Modelo AR(2)

El modelo autorregresivo de orden dos “AR(2)” se extiende a dos periodos anteriores en términos de las observaciones aleatorias y viene definido por

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (3.19)$$

Aplicando la propiedad $y_{t-r} = L^r y_t$ se obtiene

$$(1 - \phi_1 L - \phi_2 L^2)y_t = \varepsilon_t \quad (3.20)$$

El término “ $1 - \phi_1 L - \phi_2 L^2$ ” denominado operador polinomial de retardos, evaluando en $1 - \phi_1 L - \phi_2 L^2 = 0$ requiere que sus raíces estén fuera del círculo de la unidad para que el proceso “AR(2)” sea estacionario.

De manera análoga al modelo “AR(1)” se obtienen los estimadores media, varianza, autocovarianza, autocorrelación, para ello multipliquemos a la ecuación 3.19 por “ y_{t-r} ” y aplicando esperanzas se tiene

$$\begin{aligned} E[y_t y_{t-r}] &= E[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t) y_{t-r}] \\ &= \phi_1 E[y_{t-1} y_{t-r}] + \phi_2 E[y_{t-2} y_{t-r}] + E[\varepsilon_t y_{t-r}] \\ &= Cov(y_t y_{t-r}) \end{aligned} \quad (3.21)$$

La expresión anterior es obtenida debido a la condición de estacionariedad que satisface

$$E[y_t] = 0 \quad \forall t \quad (3.22)$$

La autocovarianza para un periodo de orden cero $r = 0$ converge a la varianza por lo tanto de la ecuación 3.21 se deduce

$$E[y_t y_t] = \gamma_0 \quad (3.23)$$

$$\begin{aligned}
E[y_t y_{t-r}] &= E[y_t y_{t-0}] \\
&= E[(y_t)^2] \\
&= \text{Var}(y_t) \\
&= \phi_1 E[y_{t-1} y_t] + \phi_2 E[y_{t-2} y_t] + E[\varepsilon_t y_t] \\
&= \gamma_0
\end{aligned} \tag{3.24}$$

Obsérvese que el término $E[\varepsilon_t y_t]$ converge a la varianza del ruido blanco es decir

$$\begin{aligned}
E[\varepsilon_t y_t] &= E[\varepsilon_t (\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] \\
&= \phi_1 E[\varepsilon_t y_{t-1}] + \phi_2 E[\varepsilon_t y_{t-2}] + E[\varepsilon_t \varepsilon_t] \\
&= 0 + 0 + \sigma_\varepsilon^2 \\
&= \sigma_\varepsilon^2
\end{aligned} \tag{3.25}$$

Con el resultado obtenido la ecuación 3.24 puede ser escrito

$$\begin{aligned}
\gamma_0 &= \phi_1 E[y_{t-1} y_t] + \phi_2 E[y_{t-2} y_t] + E[\varepsilon_t y_t] \\
&= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2
\end{aligned} \tag{3.26}$$

Finalmente la varianza para el modelo autorregresivo de orden dos esta dada por la siguiente expresión

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2 \tag{3.27}$$

Recordemos que $E[y_t y_{t-r}] = \phi_1 E[y_{t-1} y_{t-r}] + \phi_2 E[y_{t-2} y_{t-r}] + E[\varepsilon_t y_{t-r}]$ para un periodo $r > 0$ las covarianzas entre las observaciones y_{t-i}, y_{t-r} para “ $i > 0, r > 0$ ” únicamente esta relacionada por el tiempo transcurrido entre ellas esto es $|(t-i) - (t-r)| = r - i$, además $E[\varepsilon_t y_{t-r}] = 0$.

Con todo ello la autocovarianza para un periodo $r > 0$ esta dada por la ecuación

$$\gamma_r = \phi_1 \gamma_{r-1} + \phi_2 \gamma_{r-2}; \quad \forall r > 0 \quad (3.28)$$

Para el cálculo del coeficiente de autocorrelación se requiere dividir por γ_0 a la ecuación 3.28, además la varianza es constante $\forall y_t$ y esta dada por $\sqrt{\sigma_{y_t}^2} \sqrt{\sigma_{y_s}^2} = \sigma_0^2$

$$\rho_r = \phi_1 \rho_{r-1} + \phi_2 \rho_{r-2}; \quad \forall r > 0 \quad (3.29)$$

Las autocorrelaciones también son denominadas correlogramas, a través de ellos es posible calcular los parámetros ϕ_i del modelo $AR(2)$, asignando $r = 1, r = 2$ se obtiene un sistema de dos ecuaciones

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_1 \\ \rho_2 = \phi_1 \rho_1 + \phi_2 \end{cases} \quad (3.30)$$

El sistema obtenido recibe la denominación de sistema de ecuaciones de **Yule-Walker**, como atributo a las aportaciones de Yule(1927) y Walker(1931).

Resolviendo el sistema de ecuaciones 3.30 se obtiene los coeficientes ϕ_1, ϕ_2

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

3.3.3. Modelo $AR(p)$

Un modelo autorregresivo " $AR(p)$ " generaliza la sucesión de variables aleatorias para un factor " p " de retardo y viene definido por

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (3.31)$$

Aplicando la propiedad $y_{t-n} = L^n y_t$ inherente a las observaciones aleatorias temporales se deduce

$$\begin{aligned} y_t &= \phi_1 L y_t + \phi_2 L^2 y_t + \dots + \phi_p L^p y_t + \varepsilon_t \\ &= (\phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p) y_t + \varepsilon_t \end{aligned} \quad (3.32)$$

Simplificando

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = \varepsilon_t \quad (3.33)$$

La expresión " $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ " denominado operador polinomial de retardos de orden p , requiere que sus raíces L estén fuera del círculo de la unidad, para que el proceso $AR(p)$ sea estacionario.

De forma resumida un modelo " $AR(p)$ " puede ser escrito

$$\phi(L) y_t = \varepsilon_t \quad (3.34)$$

El cálculo de la esperanza, varianza, covarianza y correlación es similar a los modelos anteriores, con la única diferencia de que ahora se extiende a un factor p de retardo.

Bajo el supuesto de estacionariedad se asume que $E[y_t] = \mu$, $\forall t$, sustituyendo en 3.31 y tomando esperanzas se tiene

$$\begin{aligned}
E[y_t] &= \phi_1 E[y_{t-1}] + \phi_2 E[y_{t-2}] + \dots + \phi_p E[y_{t-p}] + E[\varepsilon_t] \\
\mu &= \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu + E[\varepsilon_t] \\
E[\varepsilon_t] &= \mu - \phi_1 \mu - \phi_2 \mu - \dots - \phi_p \mu \\
E[\varepsilon_t] &= (1 - \phi_1 - \phi_2 - \dots - \phi_p) \mu \\
\mu &= \frac{E[\varepsilon_t]}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \\
\mu &= 0
\end{aligned} \tag{3.35}$$

La igualdad anterior demuestra que la esperanza es cero para un proceso estacionario.

Para el calculo de la varianza, autocovarianza y autocorrelación para un periodo r entre las observaciones y_t, y_{t-r} consideramos la siguiente notación

$$y_t y_{t-r} = \phi_1 y_{t-1} y_{t-r} + \phi_2 y_{t-2} y_{t-r} + \dots + \phi_p y_{t-p} y_{t-r} + \varepsilon_t y_{t-r} \tag{3.36}$$

Aplicando esperanzas y covarianzas se resulta

$$\begin{aligned}
E[y_t y_{t-r}] &= E[\phi_1 y_{t-1} y_{t-r}] + E[\phi_2 y_{t-2} y_{t-r}] + \dots + E[\phi_p y_{t-p} y_{t-r}] + E[\varepsilon_t y_{t-r}] \\
&= \phi_1 E[y_{t-1} y_{t-r}] + \phi_2 E[y_{t-2} y_{t-r}] + \dots + \phi_p E[y_{t-p} y_{t-r}] + E[\varepsilon_t y_{t-r}] \\
&= \phi_1 \gamma_{r-1} + \phi_2 \gamma_{r-2} + \dots + \phi_p \gamma_{r-p} + E[\varepsilon_t y_{t-r}] \\
&= \gamma_r
\end{aligned} \tag{3.37}$$

La covarianza de orden cero $r = 0$ converge a la varianza, ademas $E[\varepsilon_t y_{t-0}] = \sigma_\varepsilon^2$ por lo tanto

$$E[y_t y_{t-0}] = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 \tag{3.38}$$

$$\begin{aligned}
E[y_t y_{t-0}] &= \gamma_0 \\
&= \phi_1 \gamma_{0-1} + \phi_2 \gamma_{0-2} + \dots + \phi_p \gamma_{0-p} + E[\varepsilon_t y_{t-0}] \\
&= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2
\end{aligned} \tag{3.39}$$

Denota el calculo de la varianza para un modelo autorregresivo de orden p .

De manera análoga la covarianza para un periodo $r > 0$ y recordando que $E[\varepsilon_t y_{t-r}] = 0$ se obtiene

$$\gamma_r = \phi_1 \gamma_{r-1} + \phi_2 \gamma_{r-2} + \dots + \phi_p \gamma_{r-p} \tag{3.40}$$

Dividiendo por γ_0 a la ecuación 3.40, obtenemos la ecuación de diferencias de orden “ p ”, relativa a las autocorrelaciones

$$\begin{aligned}
\frac{\gamma_r}{\gamma_0} &= \phi_1 \frac{\gamma_{r-1}}{\gamma_0} + \phi_2 \frac{\gamma_{r-2}}{\gamma_0} + \dots + \phi_p \frac{\gamma_{r-p}}{\gamma_0} \\
\rho_r &= \phi_1 \rho_{r-1} + \phi_2 \rho_{r-2} + \dots + \phi_p \rho_{r-p}
\end{aligned} \tag{3.41}$$

Asignando condiciones iniciales a la sucesión $\rho_0, \rho_1, \rho_2, \dots, \rho_{p-1}$ y particularizando para $r = 1, 2, \dots, p$ se obtiene el sistema de ecuaciones de Yule-Walker

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 = \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ \dots \\ \rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{cases} \tag{3.42}$$

Resolviendo el anterior sistema obtenemos los parámetros ϕ_i del modelo $AR(p)$, se debe considerar la simetría $\rho_{-p} = \rho_p$

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad (3.43)$$

3.4. Modelos de medias móviles MA

El modelo de medias móviles es aquel que explica el valor de una determinada variable en un periodo “ t ” en función de un término independiente y una sucesión de errores correspondientes a periodos precedentes

$$y_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.44)$$

Estos modelos normalmente se denotan “ $MA(q)$ ”, el factor de retardo que corresponde a q términos de error, se asume normalmente que son variables iid (variables aleatorias independientes idénticamente-distribuidas) muestreadas de una distribución normal con media cero y varianza constante σ_ε^2 $\varepsilon_t \sim N(0, \sigma^2)$

3.4.1. Modelo MA(1)

El modelo de medias móviles con factor de retardo unitario viene definido por

$$\begin{aligned} y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= (1 - \theta_1 L) \varepsilon_t \end{aligned} \quad (3.45)$$

Para el cálculo de la varianza, covarianza, correlación multiplicamos por y_{t-r} a la ecuación 3.45 y aplicando esperanzas resulta

$$\begin{aligned}
E[y_t y_{t-r}] &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-r} - \theta_1 \varepsilon_{t-r-1})] \\
&= E[(\varepsilon_t \varepsilon_{t-r} - \theta_1 \varepsilon_t \varepsilon_{t-r-1} - \theta_1 \varepsilon_{t-1} \varepsilon_{t-r} + \theta_1 \varepsilon_{t-1} \theta_1 \varepsilon_{t-r-1})] \\
&= E[(\varepsilon_t \varepsilon_{t-r}] - \theta_1 E[\varepsilon_t \varepsilon_{t-r-1}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-r}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-r-1}]
\end{aligned} \tag{3.46}$$

Para un periodo $r = 0$ y teniendo en cuenta que $E(\varepsilon_t \varepsilon_{t'}) = 0$ para $t \neq t'$

$$\begin{aligned}
E[y_t y_t] &= E[y_t^2] \\
&= E[(\varepsilon_t^2 - \theta_1 E[\varepsilon_t \varepsilon_{t-1}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_t] + \theta_1^2 E[\varepsilon_{t-1}^2]) \\
&= \sigma_\varepsilon^2 - \theta_1 0 - \theta_1 0 + \theta_1^2 \sigma_\varepsilon^2 \\
&= \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 \\
&= (1 + \theta_1^2) \sigma_\varepsilon^2
\end{aligned} \tag{3.47}$$

Finalmente la varianza para el modelo $MA(1)$ es

$$\gamma_0 = (1 + \theta_1^2) \sigma_\varepsilon^2 \tag{3.48}$$

Asignando $r = 1$ en la ecuación 3.46 se obtiene

$$\begin{aligned}
E[y_t y_{t-1}] &= E[(\varepsilon_t \varepsilon_{t-1} - \theta_1 E[\varepsilon_t \varepsilon_{t-2}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-1}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-2}]) \\
&= E[(\varepsilon_t \varepsilon_{t-1}] - \theta_1 E[\varepsilon_t \varepsilon_{t-2}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-1}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-2}] \\
&= 0 - \theta_1 0 - \theta_1 \sigma_\varepsilon^2 + \theta_1^2 0 \\
&= -\theta_1 \sigma_\varepsilon^2
\end{aligned} \tag{3.49}$$

La expresión anterior representa la covarianza, que dista entre dos observaciones en una unidad $\gamma_1 = Cov(y_t y_{t-1}) = -\theta_1 \sigma_\varepsilon^2$.

Si r es mayor al factor de retardo q " $r > q$ " no existe relación entre los términos del

error “ ε_t ” correspondientes al mismo periodo, es decir no se origina un término $E[\varepsilon_t \varepsilon_{t'}]$ donde $t = t'$.

A efecto de demostración se asigna $r = 2$ por tanto se obtiene

$$\begin{aligned}
 E[y_t y_{t-r}] &= E[(\varepsilon_t \varepsilon_{t-r}) - \theta_1 E[\varepsilon_t \varepsilon_{t-r-1}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-r}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-r-1}]] \\
 &= E[(\varepsilon_t \varepsilon_{t-2}) - \theta_1 E[\varepsilon_t \varepsilon_{t-3}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-2}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-3}]] \\
 &= 0 - \theta_1 0 - \theta_1 0 + \theta_1^2 0 \\
 &= 0
 \end{aligned} \tag{3.50}$$

Análogamente se puede inferir que para $r > q$ todos los valores correspondientes a las autocovarianzas γ_r convergen a cero es decir “ $\gamma_r = 0$ ”.

El proceso del cálculo de los coeficientes de autocorrelación, simplemente se reduce a obtener “ ρ_1 ” debido al factor de retardo $q = 1$, entonces

$$\begin{aligned}
 \rho_1 &= \frac{Cov(y_t, y_{t-1})}{\sqrt{Var(y_t)} \sqrt{Var(y_{t-1})}} \\
 &= \frac{Cov(y_t, y_{t-1})}{\sqrt{(1 + \theta_1^2) \sigma_\varepsilon^2} \sqrt{(1 + \theta_1^2) \sigma_\varepsilon^2}} \\
 &= \frac{Cov(y_t, y_{t-1})}{\sqrt{(1 + \theta_1^2)^2 \sigma_\varepsilon^4}} \\
 &= \frac{-\theta_1 \sigma_\varepsilon^2}{(1 + \theta_1^2) \sigma_\varepsilon^2} \\
 &= \frac{-\theta_1}{(1 + \theta_1^2)}
 \end{aligned} \tag{3.51}$$

Finalmente la autocorrelación de orden uno en términos de θ_1 se escribe

$$\rho_1 = \frac{-\theta_1}{(1 + \theta_1^2)} \tag{3.52}$$

Se ha demostrado que $\gamma_r = 0 \quad \forall r > q$ y como $\rho_r = \frac{\gamma_r}{\gamma_0}$ entonces para todo $r > q$ se cumple $\rho_r = 0$.

Se denomina condición de invertibilidad al término $|\theta_1| < 1$ que permite el paso de un $MA(q)$ a un $AR(\infty)$, si la ecuación $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ se expresa $y_t + \theta_1 \varepsilon_{t-1} = \varepsilon_t$ y se efectúan N sustituciones sucesivas resulta

$$\begin{aligned}
 \varepsilon_t &= y_t + \theta_1 \varepsilon_{t-1} \\
 &= y_t + \theta_1 (y_{t-1} + \theta_1 \varepsilon_{t-2}) \\
 &= \dots \\
 &= y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \dots + \theta_1^N y_{t-N} + \theta_1^{N+1} \varepsilon_{t-(N+1)}
 \end{aligned} \tag{3.53}$$

Si $|\theta_1| < 1$ a medida que N tiende a crecer, el término $\theta_1^N \varepsilon_{t-N} + \theta_1^{N+1} y_{t-(N+1)}$ convergen a cero, en consecuencia si se cumple siempre esta condición podrá expresarse así el modelo.

$$\varepsilon_t = y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \dots \tag{3.54}$$

Un modelo $MA(1)$ es siempre estacionario con independencia de valor que tenga el parámetro θ_1 , y su condición de invertibilidad es equivalente en sentido formal a la condición de estacionariedad $|\phi_1| < 1$ de un modelo $AR(1)$.

3.4.2. Modelo MA(2)

El modelo de media móvil de orden dos, denotado $MA(2)$ viene definido por

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \tag{3.55}$$

Aplicando la propiedad de retardo en p periodos $y_{t-p} = L^p y_t$ resulta

$$\begin{aligned} y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ &= \varepsilon_t - \theta_1 L \varepsilon_t - \theta_2 L^2 \varepsilon_t \\ &= (1 - \theta_1 L - \theta_2 L^2) \varepsilon_t \end{aligned} \quad (3.56)$$

En términos del operador polinomial de retardos denotado $\theta(L) = 1 - \theta_1 L - \theta_2 L^2$, el modelo $MA(2)$ se expresa

$$y_t = \theta(L) \varepsilon_t \quad (3.57)$$

Evaluando la esperanza a la ecuación 3.57 obtenemos $E[y_t] = \theta(L)E[\varepsilon_t]$, por lo que se concluye que $E[y_t] = 0 \quad \forall t$.

Para el cálculo de las autocovarianzas, autocorrelaciones de orden r , multiplicamos por y_{t-r} al modelo $MA(2)$ y aplicando esperanzas resulta

$$E[y_t y_{t-r}] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-r} - \theta_1 \varepsilon_{t-r-1} - \theta_2 \varepsilon_{t-r-2})] \quad (3.58)$$

$$\begin{aligned} &= E[\varepsilon_t \varepsilon_{t-r}] - \theta_1 E[\varepsilon_t \varepsilon_{t-r-1}] - \theta_2 E[\varepsilon_t \varepsilon_{t-r-2}] \\ &\quad - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-r}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-r-1}] + \theta_1 \theta_2 E[\varepsilon_{t-1} \varepsilon_{t-r-2}] \\ &\quad - \theta_2 E[\varepsilon_{t-2} \varepsilon_{t-r}] + \theta_2 \theta_1 E[\varepsilon_{t-2} \varepsilon_{t-r-1}] + \theta_2^2 E[\varepsilon_{t-2} \varepsilon_{t-r-2}] \end{aligned} \quad (3.59)$$

Evaluando para $r = 0$ en la ecuación 3.59 y aplicando la propiedad $E[\varepsilon_t \varepsilon_{t'}] = 0 \quad t \neq t'$ se obtiene

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2) \sigma_\varepsilon^2 \quad (3.60)$$

$$\begin{aligned}
\gamma_0 &= E[y_t y_{t-0}] \\
&= E[\varepsilon_t^2] + \theta_1^2 E[\varepsilon_{t-1}^2] + \theta_2^2 E[\varepsilon_{t-2}^2] \\
&= \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 + \theta_2^2 \sigma_\varepsilon^2 \\
&= (1 + \theta_1^2 + \theta_2^2) \sigma_\varepsilon^2
\end{aligned} \tag{3.61}$$

De forma análoga para $r = 1$ obtenemos

$$\begin{aligned}
\gamma_1 &= E[y_t y_{t-1}] \\
&= -\theta_1 E[\varepsilon_{t-1} \varepsilon_{t-1}] + \theta_2 \theta_1 E[\varepsilon_{t-2} \varepsilon_{t-2}] \\
&= -\theta_1 \sigma_\varepsilon^2 + \theta_2 \theta_1 \sigma_\varepsilon^2 \\
&= (-\theta_1 + \theta_1 \theta_2) \sigma_\varepsilon^2
\end{aligned} \tag{3.62}$$

Finalmente para $r = 2$

$$\begin{aligned}
\gamma_2 &= E[y_t y_{t-2}] \\
&= -\theta_2 E[\varepsilon_{t-2} \varepsilon_{t-2}] \\
&= -\theta_2 \sigma_\varepsilon^2
\end{aligned} \tag{3.63}$$

Para valores mayores a dos $r > 2$ no existe relación entre los términos de error “ ε_t ” es decir $E[\varepsilon_t \varepsilon_{t'}] = 0$ siempre será cero al no existir $t = t'$.

Los términos correspondientes a los coeficientes de autocorrelación se obtienen a partir de la expresión $\rho_i = \frac{\gamma_i}{\gamma_0}$ para un periodo unitario

$$\rho_1 = \frac{(-\theta_1 + \theta_1 \theta_2)}{(1 + \theta_1^2 + \theta_2^2)} \tag{3.64}$$

$$\begin{aligned}
\rho_1 &= \frac{\gamma_1}{\gamma_0} \\
&= \frac{\text{Cov}(y_t, y_{t-1})}{\sqrt{\text{Var}(y_t)}\sqrt{\text{Var}(y_{t-1})}} \\
&= \frac{(-\theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2}{\sqrt{(1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2}\sqrt{(1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2}} \\
&= \frac{(-\theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2}{\sqrt{(1 + \theta_1^2 + \theta_2^2)^2\sigma_\varepsilon^4}} \\
&= \frac{(-\theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2}{(1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2} \\
&= \frac{(-\theta_1 + \theta_1\theta_2)}{(1 + \theta_1^2 + \theta_2^2)}
\end{aligned} \tag{3.65}$$

Debido a que la varianza siempre es constante para todo tiempo t

$$\begin{aligned}
\rho_2 &= \frac{\gamma_2}{\gamma_0} \\
&= \frac{-\theta_2\sigma_\varepsilon^2}{(1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2} \\
&= \frac{-\theta_2}{(1 + \theta_1^2 + \theta_2^2)}
\end{aligned} \tag{3.66}$$

$$\rho_r = 0 \quad \forall r > 0 \tag{3.67}$$

Los periodos de tiempo entre las observaciones originados por r deben ser menores o iguales al factor de retardo q , $r \leq q$ de tal forma que se pueda deducir todas las expresiones anteriores. Para que el proceso sea invertible se requiere que las raíces “ L ” del polinomio característico $\theta(L) = 1 - \theta_1L - \theta_2L^2$ estén fuera del círculo de la unidad.

3.4.3. Modelo MA(q)

El modelo de medias móviles de orden “ q ”, denotado $MA(q)$ se define de la siguiente forma

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3.68)$$

Los coeficientes $\theta_1, \theta_2, \dots, \theta_q$ van precedidos por el signo negativo solamente por conveniencia de notación, generalmente contienen signos positivos. Aplicando la propiedad de retardo $y_{t-p} = L^p y_t$ a la ecuación 3.68 se tiene

$$\begin{aligned} y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\ &= \varepsilon_t - \theta_1 L \varepsilon_t - \theta_2 L^2 \varepsilon_t - \dots - \theta_q L^q \varepsilon_t \\ &= (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t \\ &= \theta(L) \varepsilon_t \end{aligned} \quad (3.69)$$

De la expresión 3.69, en notación compacta el modelo $MA(q)$ se puede escribir $y_t = \theta(L) \varepsilon_t$, para que el modelo sea invertible se requiere que las raíces de la ecuación polinomial $1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q = 0$ caigan fuera del círculo de la unidad, aplicando la media aritmética y teniendo en cuenta que $E[\varepsilon_t] = 0$ se deduce

$$\begin{aligned} E[y_t] &= E[\theta(L) \varepsilon_t] \\ &= \theta(L) E[\varepsilon_t] \\ &= 0 \end{aligned} \quad (3.70)$$

Se concluye que el valor de la media es cero cualesquiera que sea el valor de θ_i .

La obtención de las autocovarianzas y autocorrelaciones para intervalos de tiempo r entre las observaciones y_t, y_{t-r} , se calculan al multiplicar y_{t-r} al modelo $MA(q)$ y aplicando esperanzas resulta

$$E[y_t y_{t-r}] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})(\varepsilon_{t-r} - \theta_1 \varepsilon_{t-r-1} - \dots - \theta_q \varepsilon_{t-r-q})] \quad (3.71)$$

Asignando $r = 0$ en la anterior ecuación y observando el comportamiento del estimador γ_0 del modelo $MA(2)$ se logra generalizar dicha propiedad a un modelo $MA(q)$, por lo tanto

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_\varepsilon^2 \quad (3.72)$$

Generalizando las autocovarianzas γ_r para valores $r = 1, 2, \dots, q$ resulta

$$\gamma_r = \begin{cases} (-\theta_r + \theta_1 \theta_{r+1} + \dots + \theta_{q-r} \theta_q) \sigma_\varepsilon^2 & r = 1, 2, \dots, q \\ 0 & \text{si } r > q \end{cases} \quad (3.73)$$

Se observa un corte brusco en q debido a que después de ese retardo su valor es cero, el mismo fenómeno presenta los coeficientes de autocorrelación que vendrán dados por

$$\rho_r = \frac{-\theta_r + \theta_1 \theta_{r+1} + \dots + \theta_{q-r} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad r = 1, 2, \dots, q \quad (3.74)$$

$$\rho_r = 0 \quad r > q \quad (3.75)$$

3.5. Modelos mixtos autorregresivos-medias móviles ARMA

Los modelos mixtos autorregresivos-medias móviles resultan de la unión de los modelos $AR(p)$ y $MA(q)$, vienen definidos de la siguiente forma

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3.76)$$

Expresando en términos de los operadores polinomiales de retardos el modelo queda expresado en forma compacta del siguiente modo

$$\begin{aligned} y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ y_t - \phi_1 L y_t - \dots - \phi_p L^p y_t &= \varepsilon_t - \theta_1 L \varepsilon_t - \dots - \theta_q L^q \varepsilon_t \\ (1 - \phi_1 L - \dots - \phi_p L^p) y_t &= (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t \\ \phi(L) y_t &= \theta(L) \varepsilon_t \end{aligned} \quad (3.77)$$

Para que el modelo sea estacionario se requiere que las raíces “ L ” de la ecuación polinomial $1 - \phi_1 L - \dots - \phi_p L^p = 0$ estén fuera del círculo de la unidad.

3.5.1. Modelo ARMA(1,1)

Un modelo autorregresivo-media móvil viene dado por la expresión

$$y_t = \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (3.78)$$

Para el cálculo de las autocovarianzas y autocorrelaciones multiplicamos a la ecuación 3.78 por y_{t-r} y aplicamos esperanzas

$$\begin{aligned} E[y_t y_{t-r}] &= E[\phi_1 y_{t-1} y_{t-r} + \varepsilon_t y_{t-r} - \theta_1 \varepsilon_{t-1} y_{t-r}] \\ &= \phi_1 E[y_{t-1} y_{t-r}] + E[\varepsilon_t y_{t-r}] - \theta_1 E[\varepsilon_{t-1} y_{t-r}] \\ &= \gamma_r \end{aligned} \quad (3.79)$$

Evaluando para $r = 0$ obtenemos la siguiente expresión

$$\gamma_0 = \phi_1 E[y_{t-1}y_t] + E[\varepsilon_t y_t] - \theta_1 E[\varepsilon_{t-1}y_t] \quad (3.80)$$

Evaluamos a que expresiones convergen $E[\varepsilon_t y_t]$ y $E[\varepsilon_{t-1}y_t]$

$$\begin{aligned} E[\varepsilon_t y_t] &= E[\varepsilon_t(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] \\ &= \phi_1 E[\varepsilon_t y_{t-1}] + E[\varepsilon_t \varepsilon_t] - \theta_1 E[\varepsilon_t \varepsilon_{t-1}] \\ &= \sigma_\varepsilon^2 \end{aligned} \quad (3.81)$$

$$\begin{aligned} E[\varepsilon_{t-1} y_t] &= E[\varepsilon_{t-1}(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] \\ &= \phi_1 E[\varepsilon_{t-1} y_{t-1}] + E[\varepsilon_{t-1} \varepsilon_t] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-1}] \\ &= \phi_1 \sigma_\varepsilon^2 + 0 - \theta_1 \sigma_\varepsilon^2 \\ &= (\phi_1 - \theta_1) \sigma_\varepsilon^2 \end{aligned} \quad (3.82)$$

Reemplazando los valores hallados en 3.82 y 3.81 en la ecuación 3.80 finalmente obtenemos

$$\begin{aligned} \gamma_0 &= \phi_1 E[y_{t-1}y_t] + E[\varepsilon_t y_t] - \theta_1 E[\varepsilon_{t-1}y_t] \\ &= \phi_1 Cov(y_{t-1}, y_t) + \sigma_\varepsilon^2 - \theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2 \\ &= \phi_1 \gamma_1 + \sigma_\varepsilon^2 - \theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2 \end{aligned} \quad (3.83)$$

Asignando $r = 1$ en la ecuación 3.79 resulta

$$\begin{aligned}
\gamma_1 &= \phi_1 E[y_{t-1}y_{t-1}] + E[\varepsilon_t y_{t-1}] - \theta_1 E[\varepsilon_{t-1} y_{t-1}] \\
&= \phi_1 E[y_{t-1}y_{t-1}] + E[\varepsilon_t y_{t-1}] - \theta_1 E[\varepsilon_{t-1} y_{t-1}] \\
&= \phi_1 E[y_{t-1}y_{t-1}] - \theta_1 \sigma_\varepsilon^2
\end{aligned} \tag{3.84}$$

De la anterior expresión 3.84 el término $E[y_{t-1}y_{t-1}]$ es equivalente a $E[y_t y_t]$ cuyo valor ya fue calculado y converge a γ_0 es decir $\gamma_0 = E[y_{t-1}y_{t-1}]$ por lo tanto se deduce

$$\gamma_1 = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \tag{3.85}$$

Para $r = 2$ en la ecuación 3.79 se obtiene

$$\begin{aligned}
\gamma_2 &= E[y_t y_{t-2}] \\
&= \phi_1 E[y_{t-1} y_{t-2}] + E[\varepsilon_t y_{t-2}] - \theta_1 E[\varepsilon_{t-1} y_{t-2}] \\
&= \phi_1 E[y_{t-1} y_{t-2}] + E[\varepsilon_t y_{t-2}] - \theta_1 E[\varepsilon_{t-1} y_{t-2}] \\
&= \phi_1 \gamma_1 + 0 - \theta_1 0 \\
&= \phi_1 \gamma_1
\end{aligned} \tag{3.86}$$

Generalizando para periodos de tiempo mayores a uno $r > 1$ las autocovarianzas se expresan en términos recursivos es decir

$$\gamma_r = \phi_1 \gamma_{r-1} \tag{3.87}$$

Sustituyendo el valor de γ_1 en γ_0 se obtiene finalmente la autocovarianza de orden cero que se escribe

$$\begin{aligned}
\gamma_0 &= \phi_1 \gamma_1 + \sigma_\varepsilon^2 - \theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2 \\
&= \phi_1(\phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2) + \sigma_\varepsilon^2 - \theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2 \\
&= \phi_1^2 \gamma_0 - \phi_1 \theta_1 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 - \theta_1 \phi_1 \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 \\
&= \frac{(1 - 2\phi_1 \theta_1 + \theta_1^2)}{(1 - \phi_1^2)} \sigma_\varepsilon^2
\end{aligned} \tag{3.88}$$

La ecuación anterior 3.88 escrito en términos no recursivos lo sustituimos en γ_1 y realizando operaciones se resulta

$$\begin{aligned}
\gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \\
&= \phi_1 \left(\frac{(1 - 2\phi_1 \theta_1 + \theta_1^2)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \right) - \theta_1 \sigma_\varepsilon^2 \\
&= \frac{\phi_1 - 2\phi_1^2 \theta_1 + \phi_1 \theta_1^2 - \theta_1 + \phi_1^2 \theta_1}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \\
&= \frac{\phi_1 - \phi_1^2 \theta_1 + \phi_1 \theta_1^2 - \theta_1}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \\
&= \frac{(\phi_1 - \theta_1) - \phi_1 \theta_1 (\phi_1 - \theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \\
&= \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2
\end{aligned} \tag{3.89}$$

Dividiendo γ_1 por γ_0 obtenemos la el coeficiente de autocorrelación ρ_1

$$\begin{aligned}
\rho_1 &= \frac{\gamma_1}{\gamma_0} \\
&= \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \\
&= \frac{(1 - 2\phi_1 \theta_1 + \theta_1^2)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \\
&= \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{(1 - 2\phi_1 \theta_1 + \theta_1^2)}
\end{aligned} \tag{3.90}$$

De manera análoga para intervalos de tiempo mayores a uno $r > 1$ se deduce

$$\begin{aligned}
\rho_r &= \frac{\gamma_r}{\gamma_0} \\
&= \frac{\phi_1 \gamma_{r-1}}{\gamma_0} \\
&= \phi_1 \frac{\gamma_{r-1}}{\gamma_0} \\
&= \phi_1 \rho_{r-1}
\end{aligned} \tag{3.91}$$

Los coeficientes de autocorrelación de los modelos $ARMA(1,1)$ decrecen siempre en valor absoluto, tomando como condición inicial ρ_1 .

3.5.2. Modelo $ARMA(p,q)$

El modelo $ARMA$ expresado en términos autorregresivos y medias móviles, con factores de retardo p, q viene definido

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \tag{3.92}$$

Para hallar todas las autocovarianzas y autocorrelaciones para diferentes periodos r entre las observaciones y_t, y_{t-r} se multiplica por y_{t-r} a la ecuación anterior 3.92

$$y_t y_{t-r} - \phi_1 y_{t-1} y_{t-r} - \dots - \phi_p y_{t-p} y_{t-r} = \varepsilon_t y_{t-r} - \theta_1 \varepsilon_{t-1} y_{t-r} - \dots - \theta_q \varepsilon_{t-q} y_{t-r} \tag{3.93}$$

Aplicando esperanzas a ecuación 3.93 y recordando que el ruido blanco “ ε_t ” no incide en observaciones “ $y_{t'}$ ” para $t < t'$, en otros términos $E[\varepsilon_t y_{t'}] = 0 \quad t > t'$ por lo tanto

$$E[\varepsilon_t y_{t-r}] - \theta_1 E[\varepsilon_{t-1} y_{t-r}] - \dots - \theta_q E[\varepsilon_{t-q} y_{t-r}] = 0 \tag{3.94}$$

Resolviendo

$$\begin{aligned} 0 &= E[y_t y_{t-r}] - \phi_1 E[y_{t-1} y_{t-r}] - \dots - \phi_p E[y_{t-p} y_{t-r}] \\ &= \gamma_r - \phi_1 \gamma_{r-1} - \dots - \phi_p \gamma_{r-p} \end{aligned} \quad (3.95)$$

Se observa que las autocovarianzas se obtienen en términos recursivos para valores $r > q$

$$\gamma_r = \phi_1 \gamma_{r-1} + \dots + \phi_p \gamma_{r-p} \quad r > q \quad (3.96)$$

De forma análoga obtenemos las autocorrelaciones a partir de la anterior ecuación 3.96

$$\begin{aligned} \rho_r &= \frac{\gamma_r}{\gamma_0} \\ &= \frac{\phi_1 \gamma_{r-1} + \dots + \phi_p \gamma_{r-p}}{\gamma_0} \\ &= \phi_1 \frac{\gamma_{r-1}}{\gamma_0} + \dots + \phi_p \frac{\gamma_{r-p}}{\gamma_0} \\ &= \phi_1 \rho_{r-1} + \dots + \phi_p \rho_{r-p} \end{aligned} \quad (3.97)$$

Finalmente resulta

$$\rho_r = \phi_1 \rho_{r-1} + \dots + \phi_p \rho_{r-p} \quad r > q \quad (3.98)$$

La ecuación de diferencias obtenida permite generar los coeficientes de autocorrelación para $r > q$, en la determinación de los “ q ” primeros valores “ ρ_r ” interviene la parte de medias móviles del modelo.

3.6. Modelo ARIMA

Una buena parte de las series económicas se deben considerar generadas por procesos no estacionarios, desde esta perspectiva se considera al modelo $AR(1)$ cuyo coeficiente $\phi_1 = 1$

$$y_t = y_{t-1} + \varepsilon_t \quad (3.99)$$

A este modelo se le denomina “paseo aleatorio o random walk”, cuando el paseo se inicia en un pasado remoto N , realizando sustituciones sucesivas se puede expresar así

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ &= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= \dots \\ &= y_{t-N} + \sum_{j=0}^{N-1} \varepsilon_{t-j} \end{aligned} \quad (3.100)$$

A medida que N tiende a infinito la observación “ y_{t-N} ” pierde relevancia, por lo tanto resulta

$$y_t = \sum_{j=0}^{\infty} \varepsilon_{t-j} \quad (3.101)$$

La varianza del proceso es infinita en consecuencia el modelo es no estacionario, suponiendo que el proceso se inicia en $-N$ la varianza para el periodo “ t ” viene dada por la siguiente expresión

$$\gamma_{0t} = \sigma_{\varepsilon_{t-0}}^2 + \sigma_{\varepsilon_{t-1}}^2 + \dots + \sigma_{\varepsilon_{t-N}}^2 \quad (3.102)$$

Al tomar la varianza valores distintos para cada tiempo t tampoco es estacionario, una gran mayoría de los eventos económicos, demográficos, meteorológicos son de naturaleza no estacionaria, pero pueden ser transformados a modelos estacionarios de la siguiente forma

$$\begin{aligned} w_t &= y_t - y_{t-1} \\ &= \Delta y_t \\ &= \varepsilon_t \end{aligned} \quad (3.103)$$

La expresión anterior resulta de aplicar diferencias de primer orden al modelo original, ahora consideremos el caso inverso es decir como obtener y_t a partir de w_t entonces

$$y_t = w_t + y_{t-1} \quad (3.104)$$

Realizando sustituciones sucesivas en términos de w_t en la ecuación 3.104 obtenemos

$$\begin{aligned} y_t &= w_t + y_{t-1} \\ &= w_t + w_{t-1} + y_{t-2} \\ &= w_t + w_{t-1} + w_{t-2} + y_{t-3} \\ &= \dots \\ &= w_t + w_{t-1} + w_{t-2} + w_{t-3} + \dots \end{aligned} \quad (3.105)$$

Por lo tanto el proceso y_t se obtiene sumando o integrando el proceso w_t , por esta razón se dice que el “*paseo aleatorio*” pertenece a la clase de modelos integrados.

Los modelos integrados son aquellos que se pueden obtener mediante suma o integración de un proceso estacionario, a estos modelos se los denomina también modelos no estacionarios homogéneos estudiados por Tintner (1940), Tintner y Rao (1963), Yunglom(1955), recientemente por Box y Jenkins (1976) quienes contribuyeron a su divulgación.

A un proceso integrado y_t se le denomina proceso $ARIMA(p, d, q)$ si tomando primeras diferencias d – veces se obtienen un proceso estacionario w_t del tipo $ARMA(p, q)$.

La vocal I de $ARIMA$ significa integrado, con los criterios mencionados anteriormente se tiene lo siguiente

$$\begin{aligned} w_t &= \Delta y_t \\ &= (1 - L)^d y_t \end{aligned} \quad (3.106)$$

Realizando un ejemplo para un caso particular de integración $d = 2$ se tiene

$$\begin{aligned} (1 - L)^2 y_t &= (1 - 2L + L^2) y_t \\ &= y_t - 2Ly_t + L^2 y_t \\ &= y_t - 2y_{t-1} + y_{t-2} \\ &= [y_t - y_{t-1}] - [y_{t-1} - y_{t-2}] \end{aligned} \quad (3.107)$$

Aplicando esta escala de integración

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) w_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t \quad (3.108)$$

Sustituyendo w_t en la ecuación 3.108 resulta

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d y_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t \quad (3.109)$$

En términos de los operadores polinomiales de retardo

$$\phi(L)(1 - L)^d y_t = \theta(L) \varepsilon_t \quad (3.110)$$

Los modelos ARIMA constituyen una clase particular de procesos no estacionarios, sin embargo en muchos casos son suficientes para representar el comportamiento de las series no estacionarias

3.6.1. Identificación del modelo ARIMA

La fase de identificación del modelo $ARIMA(p, d, q)$ corresponde a determinar numéricamente los parámetros $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, d, \lambda, \mu, \sigma_\varepsilon^2$ de un proceso estocástico expresado en términos de la ecuación 3.111, para hallar un proceso w_1, w_2, \dots, w_r estacionario

$$\phi_p(L)(1 - L)^d (y_t^\lambda - \mu) = \theta_q(L) \varepsilon_t \quad (3.111)$$

La fase de identificación comprende

- Se procede a efectuar un análisis de estacionariedad a la serie, en el caso de que se trate de una serie no estacionaria se aplica el proceso de integración al objeto de convertirla en estacionaria, es decir dado $w_t = (1 - L)^d y_t^\lambda$ se eligen valores d, λ de forma que sea estacionaria
- Hallar los factores de retardo p, q de la parte autorregresiva $AR(p)$ y medias móviles $MA(q)$ del proceso ARMA que se considere hay podido generar la serie estacionaria w_t

- Realizar pronósticos en términos probabilísticos de valores futuros de la variable, en el caso de existir analogías entre los valores pronosticados y los que se observen realmente se deberá cuestionar la validez del modelo identificado

3.6.2. Instrumentos estadísticos de identificación

Los instrumentos básicos de identificación de los modelos estacionarios son la función de autocorrelación FAC y la función de autocorrelación parcial $FACP$ que son instrumentos desarrollados para la identificación de modelos $ARMA$.

Función de autocorrelación FAC

Dado una serie estacionaria w_t , $\phi_p(L)w_t = \theta(L)_q$ de tamaño N se evalúa el coeficiente de autocorrelación muestral de orden r mediante el siguiente estimador

$$\rho_r = \frac{\sum_{t=r+1}^N (w_t - \bar{w})(w_{t-r} - \bar{w})}{\sum_{t=1}^N (w_t - \bar{w})^2} \quad (3.112)$$

A medida que crece el orden de el retardo r disminuye el número de sumandos, así para r_{N-1} solamente se dispondrá de un solo sumando en el numerador y para r_N ya no se podría calcular, surge el inconveniente de perder eficiencia en la estimación por ello se recomienda calcular coeficientes de autocorrelación ρ_r para valores de r superiores a un tercio o un cuarto del tamaño de la muestra.

Función de autocorrelación parcial FACP

En un modelo $AR(p)$ la función de autocorrelación FAC empieza a decrecer a partir de un determinado retardo pero nunca se hace cero, por esta razón es difícil discriminar a la vista del correlograma el orden de un proceso autorregresivo, la función de autocorrelación parcial $FACP$ permitiría identificar el orden del modelo autorregresivo.

Recordemos que los coeficientes ϕ_i de un modelo $AR(p)$ se pueden calcular a través de las ecuaciones de *Yule – Walker* en función de los coeficientes de autocorrelación

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad (3.113)$$

Resolviendo el sistema para $AR(1)$, $AR(2)$, \dots , $AR(p)$, se obtendrán los siguientes coeficientes

$$\begin{array}{l} AR(1) \quad \phi_{11} \\ AR(2) \quad \phi_{21} \quad \phi_{22} \\ \vdots \\ AR(p) \quad \phi_{p1} \quad \phi_{p2} \quad \dots \quad \phi_{pp} \end{array} \quad (3.114)$$

La notación ϕ_{ij} es relativa a identificar el orden del modelo $i = 1, 2, \dots, p$ y j el coeficiente del término $\phi_j y_{t-j}$ en un modelo de orden i .

Tomando el último coeficiente de cada uno de los procesos $AR(1)$, $AR(2)$, \dots , $AR(p)$, se obtiene la función de autocorrelación parcial $FACP$

$$\phi_{11}, \phi_{22}, \phi_{33}, \dots, \phi_{pp} \quad (3.115)$$

Identificación de un modelo $ARMA(p, q)$

Para la identificación del modelo generador $ARMA(p, q)$ los instrumentos básicos son FAC , $FACP$. El proceso de identificación consiste en comparar el comportamiento de estos estadísticos con funciones de autocorrelación y autocorrelación parcial correspondientes a distintos modelos teóricos con los que puedan guardar similitud, teniendo en cuenta que nunca cabe esperar una similitud perfecta debido a errores de muestreo

Modelos $AR(p)$

En un modelo $AR(p)$ los coeficientes de autocorrelación presentan un decrecimiento rápido que pueden ser geométrico puro, geométrico con alternancia de signos, sinusoidal, dependerá del signo y las características de las raíces del proceso generador.

En la figura 3.1 se presenta la función de autocorrelación teórica FAC y la función de autocorrelación parcial $FACP$ para distintos procesos $AR(1)$ y $AR(2)$

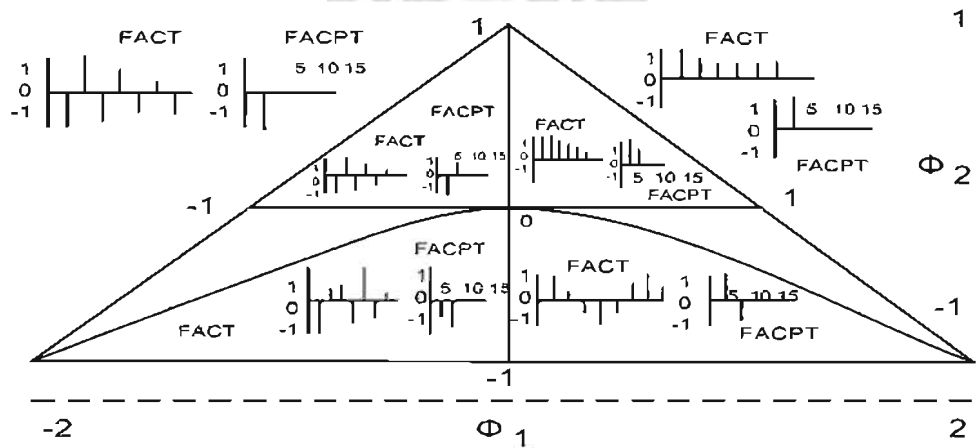


Figura 3.1: FAC y FACP en los modelos $AR(1)$ y $AR(2)$
Fuente: Ezequiel U. Peiró A.

Un modelo $AR(2)$ será estacionario si los coeficientes ϕ_1, ϕ_2 corresponden al lugar geométrico acotado por las ecuaciones 3.116 que forman un triángulo

$$\begin{aligned}
 \phi_2 &= 1 - \phi_1 \\
 \phi_2 &= 1 + \phi_1 \\
 \phi_2 &= -1
 \end{aligned}
 \tag{3.116}$$

La región que aparece debajo del arco de la parábola corresponde a modelos cuya ecuación en diferencias tiene raíces imaginarias, un modelo $AR(1)$ es un caso particular de un $AR(2)$ cuando $\phi_2 = 0$.

Modelos $MA(q)$

En un modelo $MA(q)$ la FAC toma el valor de cero para retardos superiores a q , la función de autocorrelación parcial $FACP$ decrece en forma rápida, sin llegar a anularse nunca es un comportamiento recíproco al modelo $AR(p)$.

En la figura 3.2 se observa las funciones FAC , $FACP$ de seis estructuras de modelos $MA(1)$, $MA(2)$ cuyos coeficientes θ_1, θ_2 corresponden a la región de invertibilidad acotados por las rectas

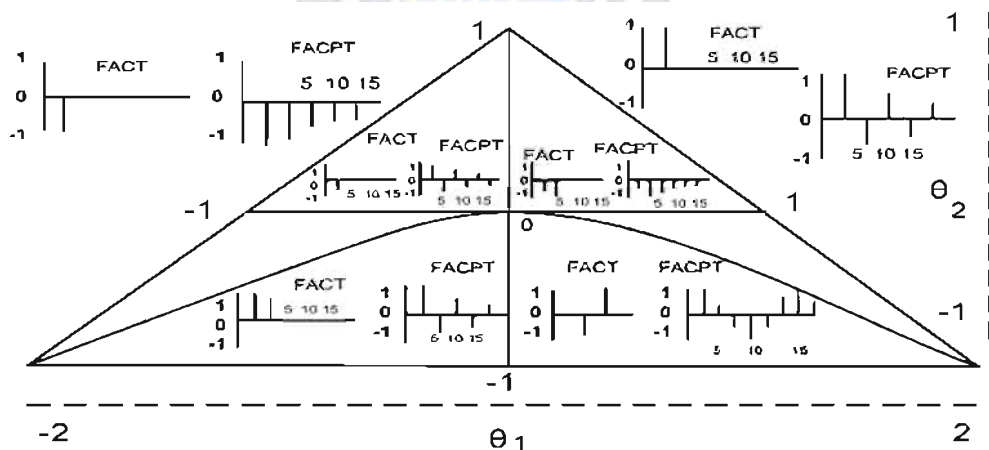


Figura 3.2: FAC y $FACP$ en los modelos $MA(1)$ y $MA(2)$

Fuente: Ezequiel U. Peiró A.

$$\begin{aligned}
 \theta_2 &= 1 - \theta_1 \\
 \theta_2 &= 1 + \theta_1 \\
 \theta_2 &= -1
 \end{aligned}
 \tag{3.117}$$

La región que aparece debajo de la parábola corresponde a modelos con raíces imaginarias.

Modelos $ARMA(p,q)$

En un modelo $ARMA(p,q)$ los coeficientes ρ_1, \dots, ρ_q asociados a FAC , y ρ_1, \dots, ρ_p asociados a $FACP$ no tienen un patrón fijo de comportamiento y decrecen según una

MODELO	FAC	FACP
AR(p)	Decrecimiento rápido de tipo geométrico puro y geométrico con alternancia de signos, sinusoidal o mezcla de varios tipos	Se anula para retardos superiores a p
MA(q)	Se anula para retardos superiores a q	decrecimiento rápido de tipo exponencial y/o sinusoidal.
ARMA(p,q)	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas

Cuadro 3.1: FAC y FACP en los modelos AR(p),MA(q),ARMA(p,q)
Fuente: Ezequiel U. Peiró A.

mezcla de oscilaciones sinusoidales y exponenciales.

En la figura 3.3 se representan las funciones *FAC* y *FACP* de seis modelos *ARMA(1,1)* estacionarios e invertibles, la región conjunta de estacionariedad e invertibilidad esta definida por todas las combinaciones situadas en el interior del cuadrado.

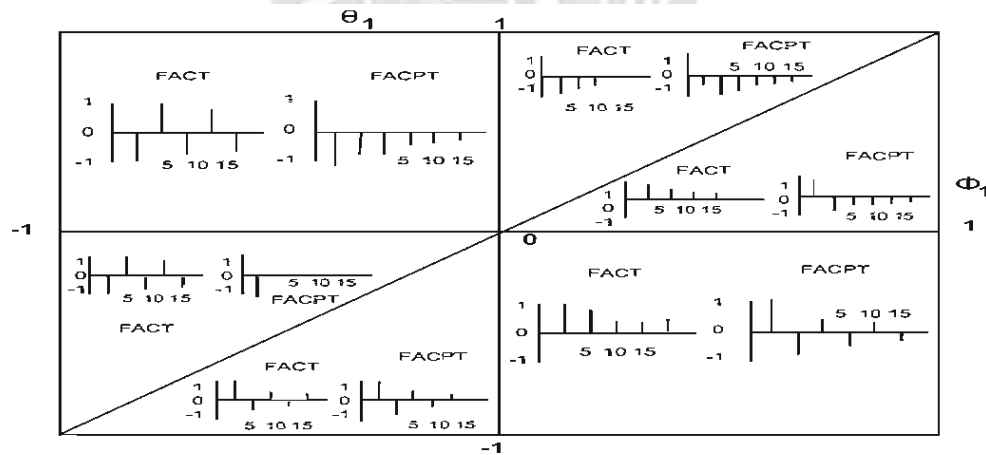


Figura 3.3: FAC y FACP en los modelos ARMA(1,1)
Fuente: Ezequiel U. Peiró A.

El cuadro 3.1 contiene un resumen del comportamiento de las funciones *FAC*, *FACP* para los modelos *AR(p)*, *MA(q)* y *ARMA(p,q)*.

Capítulo 4

Construcción de la Red Neuronal

4.1. Introducción

En las secciones correspondientes a este capítulo se elabora la teoría necesaria que permitirá aplicar la red neuronal a un proceso estocástico, análogamente se obtendrán los modelos estadísticos $ARIMA(p, d, q)$ que serán un estimador en el proceso de contraste de resultados respecto de los datos reales y los modelos basados en redes neuronales artificiales.

Se plantean posibles arquitecturas de red neuronal y el criterio de asociación de la información en la búsqueda de la topología adecuada que permita realizar un ajuste óptimo a la serie estudiada.

4.2. Muestra aleatoria

Sea una muestra aleatoria de tamaño n correspondiente a un proceso estocástico es decir una sucesión de variables temporales correspondientes a eventos: Climatológicos, económicos, financieros, etc.

$$y_1, y_2, \dots, y_n \tag{4.1}$$

El comportamiento natural de la observaciones dado un valor origen y_i , es un fenómeno que puede relacionarse a una dependencia de valores precedentes, esto es la observación y_i depende en cierto grado de la observación y_{i-1} considerando un factor meramente aleatorio ello implica que puede existir otros factores que incidan en la relación.

A continuación se describen los tipos de relaciones que pueden existir entre las variable muestreadas.

- y_i no tenga en absoluto alguna relación con y_{i-1} .
- y_i tiene un grado de relación superfluo con y_{i-1} .
- y_i depende absolutamente de y_{i-1} .

Los eventos de series temporales generalmente se comportan con cierto grado de dependencia absoluta, debido a que siempre existirán factores aleatorios que incidirán en el comportamiento de dichos eventos.

Las observaciones y_1, \dots, y_{n-1} ejercen un grado de impacto en la observación y_n debido a que los sucesos anteriores ejercen influencia en los eventos presentes y futuros, es necesario considerar que a medida que n tiende a crecer indefinidamente las observaciones iniciales y_1, \dots, y_r pierden relevancia al no ejercer alguna influencia significativa en la observación reciente debido a ello estas observaciones generalmente no son consideradas.

Mediante procesos de prueba y error se concluyo que una sucesión de tres observaciones $X = (x_1, x_1, \dots, x_n)$ es un criterio adecuado para asignar al conjunto de entrenamiento, vectores de dimensión R^3 ; Toda la teoría relativa a estas afirmaciones se expone en los dos siguientes acápite.

4.3. Conjunto de entrenamiento

Las redes neuronales artificiales responden a criterios *autoasociativos* en los cuales solo se necesitan datos de entrada y los criterios *heteroasociativos* donde la información de entrada necesariamente esta asociado a un valor de salida.

La heteroasociatividad será la propiedad que se utilizará en la obtención del modelo generador de las observaciones temporales basado en redes neuronales artificiales.

A continuación se deducen los criterios bajo los cuales los datos correspondientes a un proceso estocástico son transformados en un conjunto de entrenamiento de la red neuronal artificial.

Sea la muestra aleatoria y_1, y_2, \dots, y_n , la distancia entre los valores y_i, y_j $i < j$ estará medido en términos de las observaciones y_k , ello implica que pueden existir $k - \text{observaciones}$ entre los valores y_i, y_j .

En particular si $k = 0$ en la sucesión temporal y_1, y_2, \dots, y_n existe una dependencia directa es decir y_2 depende de y_1 , y_3 depende de y_2 y sucesivamente y_n depende de y_{n-1} , se puede observar esta dependencia en el siguiente cuadro 4.1

Observación	Valor deseado
(y_1)	y_2
(y_2)	y_3
(y_3)	y_4
...	...
(y_{n-1})	y_n

Cuadro 4.1: Asociación de la información, dimensión 1
Fuente: Elaboración propia

Naturalmente una dependencia de la observación actual será mejor cuanto mayor sea el número de observaciones precedentes siendo así, si $k=1$ se tiene una observación intermedia entre los valores y_i, y_j $i < j$, bajo estos criterios la sucesión 4.1 se transforma en la siguiente distribución de observaciones y valores deseados como se puede apreciar esta dependencia en el siguiente cuadro 4.2

La observación y_3 depende de las observaciones precedentes y_2 y y_1 , la observación y_4 depende de y_3 y y_2 sucesivamente y_n depende de y_{n-1} y y_{n-2} .

De forma análoga si $k = 2$ existen dos observaciones intermedias entre los valores y_i, y_j $i < j$, con estos criterios la sucesión 4.1 tiene la siguiente distribución de observaciones y valores deseados como se aprecia en el siguiente cuadro 4.3

Observación	Valor deseado
(y_1, y_2)	y_3
(y_2, y_3)	y_4
(y_3, y_4)	y_5
...	...
(y_{n-2}, y_{n-1})	y_n

Cuadro 4.2: Asociación información dimensión 2
Fuente: Elaboración propia

Observación	Valor deseado
(y_1, y_2, y_3)	y_4
(y_2, y_3, y_4)	y_5
(y_3, y_4, y_5)	y_6
...	...
$(y_{n-3}, y_{n-2}, y_{n-1})$	y_n

Cuadro 4.3: Asociación información dimensión 3
Fuente: Elaboración propia

La cantidad de observaciones entre los valores y_i, y_j $i < j$ se halla en el intervalo $0 \leq k \leq (n - 2)$ siendo n la cantidad total de observaciones a ser estudiadas, generalizando cuando k converge a $(n - 2)$, la sucesión y_1, y_2, \dots, y_{n-1} se transforma en el vector de entrenamiento para la red neuronal artificial asociado a un único valor de salida deseado y_n , esta distribución se puede observar en el siguiente cuadro 4.4

Observación	Valor deseado
$(y_1, y_2, y_3, \dots, y_{n-1})$	y_n

Cuadro 4.4: Asociación información dimensión (n-1)
Fuente: Elaboración propia

4.4. Dimension de los vectores de entrenamiento

En el límite superior " $k = n - 2$ " se vio que la sucesión y_1, y_2, \dots, y_{n-1} es el conjunto de entrenamiento de la red neuronal, de acuerdo a este criterio existe un único vector de entrada $P = (y_1, y_2, \dots, y_{n-1})$ de dimensión $(n - 1)$ que será utilizado para entrenar la red y hallar el conjunto de pesos adecuados, la desventaja surge al no poder contar con varios

vectores P de entrenamiento debido a ello la red tendrá que aprender con un único vector de entrada como consecuencia la salida pueda no converger al valor deseado.

Para una arquitectura de red neuronal multicapa el vector $P = (y_1, y_2, \dots, y_{n-1})$ de entrenamiento queda representado como se puede apreciar en la siguiente figura 4.1

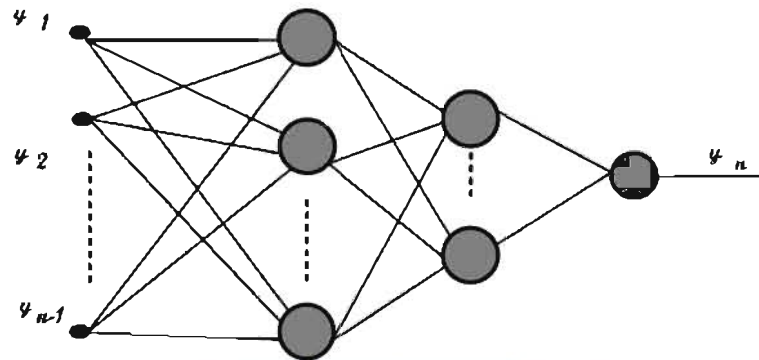


Figura 4.1: Arquitectura RNA para la sucesión $(n - 1)$
Fuente: Elaboración propia.

Cuando $k = 0$ el vector de entrenamiento $P = (y_i)$ esta formado por un único componente que se asocia a un valor deseado y_{i+1} de acuerdo a esta heteroasociación de la información la arquitectura de red neuronal artificial será la más elemental (básica), solamente se dispone de un único valor numérico en los reales como conjunto de entrenamiento asociado a un único valor de salida, se puede apreciar esta característica en la siguiente figura 4.2

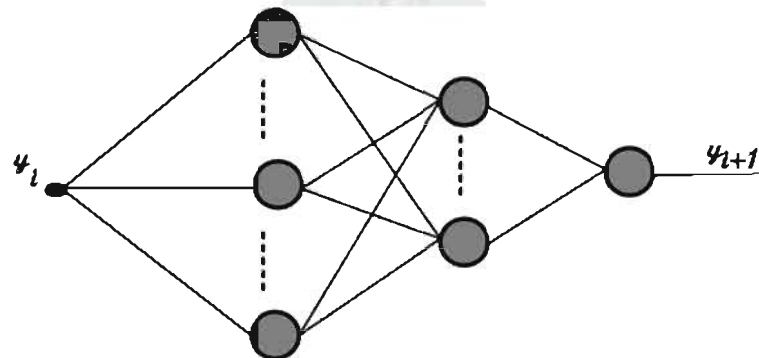


Figura 4.2: Arquitectura RNA para la sucesión (1)
Fuente: Elaboración propia.

La ventaja consiste en que se puede disponer de $(n - 1)$ vectores de entrenamiento que pueden ser utilizados para perfeccionar los pesos y hallar los valores deseados, la dificultad radica en que no se dispone de mucha información respecto al pasado el cual aportaría significativamente al las observaciones presentes y futuras.

Para un numero de observaciones intermedias $k = 2$ la serie aleatoria temporal y_1, \dots, y_n contiene los siguientes vectores de entrenamiento que corresponden al espacio R^3 y valores deseados d_i , se aprecia en el cuadro 4.5

Vectores entrenamiento P	Valor deseado d_i
(y_1, y_2, y_3)	y_4
(y_2, y_3, y_4)	y_5
(y_3, y_4, y_5)	y_6
\dots	\dots
$(y_{n-3}, y_{n-2}, y_{n-1})$	y_n

Cuadro 4.5: Vectores entrenamiento y valores deseados dimensión R^3

Fuente: Elaboración propia

La arquitectura de la red multicapa correspondiente a estas características en la asociación de la información se observa en la figura 4.3

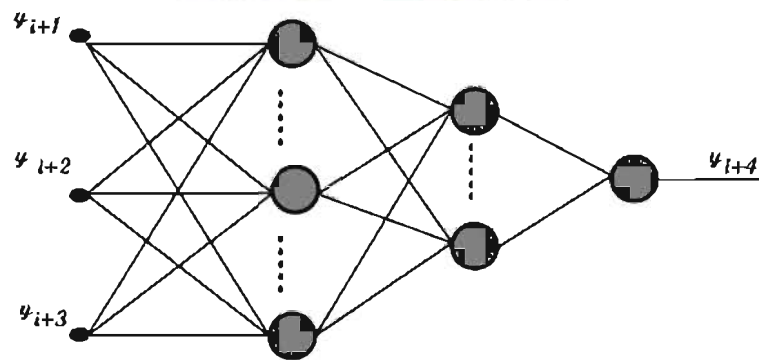


Figura 4.3: Arquitectura RNA para la sucesión (3)

Fuente: Elaboración propia

Obsérvese que a medida que aumenta la dimensión $(r - 1)$ de los vectores de entrenamiento $P = (y_1, y_2, \dots, y_{r-1})$, los valores deseados tienden a reducir y estarán formados por la sucesión $y_r, y_{r+1}, y_{r+2}, \dots, y_n$.

4.5. Diseño experimental

En el proceso de investigación y su aplicabilidad en sentido práctico, se requieren identificar claramente todos los elementos que permitirán medir el grado de precisión en la obtención de resultados.

Se realizarán una serie de entrenamientos de la red neuronal artificial con diferentes números de capas y neuronas, se trabajarán con las funciones de activación logaritmo sigmoideal y tangente sigmoideal en la búsqueda de una arquitectura optima de red que permita realizar el ajuste a la serie temporal estudiada.

4.5.1. Identificación de variables de la hipótesis

Variables independientes

- Número de capas
- Número de neuronas por capa
- Número de iteraciones (epochs)
- Error mínimo permitido (goal)
- Factor de aprendizaje (lr)

Variables dependientes

- Grado de ajuste del modelo obtenido respecto de la serie real

4.5.2. Instrumento de medida

El estimador que permitirá reflejar el grado de aproximación o ajuste del modelo obtenido mediante redes neuronales, será el error cuadrático medio, instrumento estadístico que evalúa la sumatoria de las distancias entre los valores reales y_i y los obtenidos \hat{y}_i hallando el promedio entre el numero de observaciones m .

$$ECM = \sum_{i=1}^m \frac{(y_i - \hat{y}_i)^2}{m} \quad (4.2)$$

4.5.3. Registro experimental

El proceso experimental requiere un adecuado seguimiento de las siguientes variables relativas a las redes neuronales

- Número de capas.
- Número de neuronas por capa.
- Error cuadrático medio ECM.
- Número máximo de iteraciones (epochs).
- Error mínimo a alcanzar (goal).
- Factor de aprendizaje (lr).

Análogamente en los modelos ARIMA se requiere identificar las siguientes variables

- Estacionalidad de la serie temporal.
- Función de autocorrelación FAC.
- Función de autocorrelación parcial FACP.
- Modelo ARIMA asociado.
- Criterios de información Akaike AIC y BIC.

En el proceso de pruebas se llevará un registro metódico correspondiente a las variables mencionadas las cuales deberán proveer la información necesaria que permitirán tomar decisiones respecto a los modelos estudiados.

4.6. Prototipo

El software fue desarrollado en la plataforma Matlab 7.0 presenta una interfaz gráfica que permite apreciar el proceso de ajuste a la serie temporal estudiada.

En la figura 4.4 se puede apreciar el entorno grafico que presenta diferentes funcionalidades orientadas específicamente al trabajo con redes neuronales.

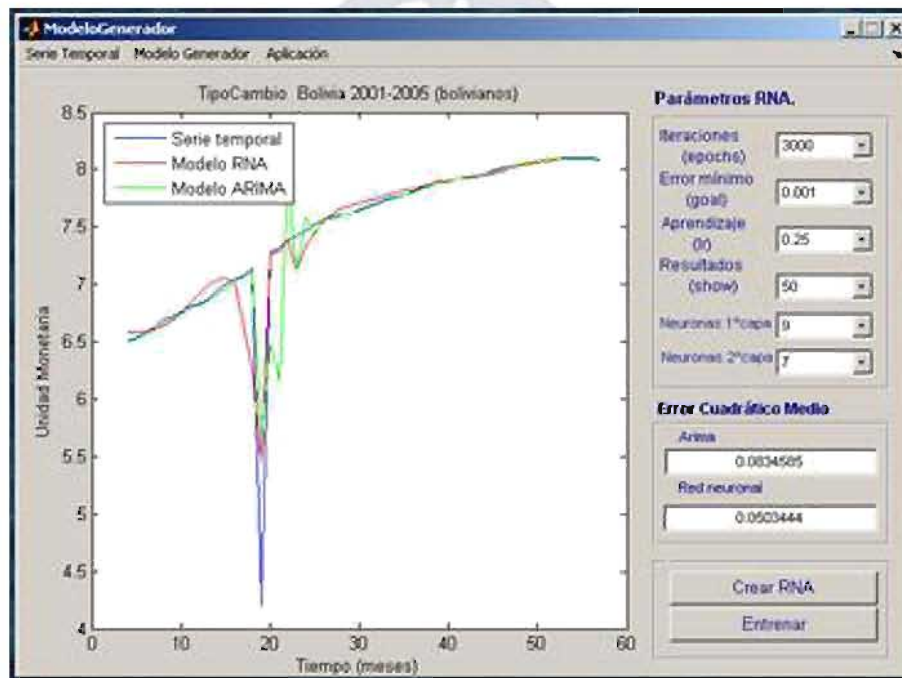


Figura 4.4: Prototipo interfaz grafica
Fuente: Elaboración propia.

4.6.1. Descripción del prototipo

Crear RNA

Permite inicializar una nueva red neuronal con pesos aleatorios correspondiente a una arquitectura de tres capas, dos capas ocultas con un número de neuronas opcional y la capa de salida con una única neurona que despliega toda la secuencia de la serie menos tres observaciones iniciales (y_1, y_2, y_3) que corresponden a un secuencia de asignación para los vectores de entrenamiento.

Entrenar

Con los parámetros asignados inicia el proceso de entrenamiento hasta alcanzar la cota mínima de error o el máximo número de iteraciones, el proceso de entrenamiento consiste en la modificación de los pesos de la red de forma que pueda clasificar correctamente los patrones de entrada.

Parámetros

Permiten controlar el numero máximo de iteraciones (epochs) que se realizará en el proceso de entrenamiento, el error mínimo deseado(goal) que querramos que alcance la red neuronal, el factor de aprendizaje(lr) que permite controlar la velocidad con que se va modificando los pesos en la red, el intervalo de visualización del error alcanzado cada cierto número de iteraciones (Show), los números de neuronas en la primera y segunda capa el valor cero implica que no contiene alguna neurona en la capa respectiva.

Error Cuadrático Medio

Permite visualizar los errores cuadráticos promedio en base al estimador $\varepsilon = \frac{\sum_{i=1}^N (d_i^k - f_i^k)^2}{N}$ de los modelos obtenidos por el proceso ARIMA y redes neuronales respecto de la serie real estudiada, se aplica el estimador ε a las salidas de la red y la serie real, de forma similar se realiza con el proceso ARIMA para obtener los errores cuadráticos base de decisión para el modelo óptimo.

Serie Temporal

Provee la opción de apertura a una nueva serie temporal para su estudio desde un explorador de Windows, en el proceso se aplica el algoritmo de asociación de la información que consiste en la obtención de los vectores de entrenamiento y valores deseados, análogamente permite inicializar todo el proceso de entrenamiento.

Modelo Generador

Contiene opciones que permiten obtener el modelo generador de la serie temporal en

estudio en base a los procesos ARIMA y Redes Neuronales Artificiales, se despliega la grafica correspondiente a los modelos obtenidos y la serie real desplegando los errores cuadráticos promedio.

Aplicación

Provee opciones de predicción a 12 observaciones futuras utilizando la información almacenada en los pesos de la Red Neuronal, para el proceso ARIMA se despliega una banda de confianza entre $\pm 1,96 * \sigma$, una otra aplicación es relativa al reconocimiento de ondas provenientes de electrocardiogramas para el caso de la fibrilación ventricular que es equivalente a un para cardiaco.

4.6.2. Algoritmos

El software desarrollado trabaja básicamente sobre tres algoritmos principales: Algoritmo de asociación de la información para el conjunto de entrenamiento, algoritmo de estandarización de la información con medio cero y varianza uno, algoritmo de aprendizaje de la red neuronal.

Algoritmo para la asociación de la información

- a.- Muestra finita y_1, y_2, \dots, y_n
- b.- Factor de asociación $k < n$ con k numero entero
- c.- Obtener la secuencia de cada k-elementos
- d.- Repetir el paso 3 hasta el tamaño máximo de la serie
- e.- Decomponer la secuencia obtenida en los vectores de entrenamiento y valores deseados

Algoritmo para la estandarización de la información

- a.- Conjuntos de entrenamiento y valores deseados

- b.- Obtener la media estadística μ para cada conjunto de entrada
- c.- Obtener las varianzas σ^2 y desviaciones estándar σ para los datos de entrada
- d.- Normalizar las observaciones según $y_i \sim N(\mu, \sigma)$, $z = \frac{y_i - \mu}{\sigma}$

Algoritmo para el aprendizaje de la red neuronal

Algoritmo Backpropagation

- a.- Se define el número de capas y neuronas de la red asociados a un conjunto de pesos (W) aleatorios
- b.- Se presenta a la red los vectores de entrenamiento $X = (x_1, x_1, \dots, x_n)$ y valores deseados d_i
- c.- Evaluar la suma ponderada o producto escalar en la capa j -ésima e i -ésima unidad sigmoideal $S_i^j = X^{j-1} W_i^j$
- d.- Aplicar las funciones de transferencia o activación en toda capa y neurona $f_i^j(S_i^j)$
- e.- Las salidas f_i^j de la capa j son los nuevos vectores de entrada para la capa $j + 1$
- f.- Evaluar la sensibilidad del error cuadrático en la capa de salida $\delta^k = (d - f)f(1 - f)$
- g.- Actualizar los pesos en la ultima capa según $W^k = W^k + \eta^k (d - f)f(1 - f)X^{k-1}$
- h.- Evaluar la sensibilidad del error cuadrático $\delta_i^j = f_i^j(1 - f_i^j) \sum_{q=1}^{m_{j+1}} \delta_q^{j+1} w_{iq}^{j+1}$ en las capas ocultas
- i.- Actualizar los pesos en las capas ocultas según $W_i^{(j)} = W_i^{(j)} + \eta_i^{(j)} \delta_i^{(j)} X^{(j-1)}$ es un proceso de retropropagacion del error hacia la capa de salida
- j.- Las salidas $y_i = f_i^k$ de la última capa son comparadas con los valores deseados hallando el termino de error $\varepsilon = (d_i - y_i)$
- k.- Repetir la secuencia c,d,e,f,g,h,i,j hasta que el error $\varepsilon = \frac{\sum_{i=1}^N (d_i^k - f_i^k)^2}{N}$ sea \leq al deseado

Capítulo 5

Resultados Prueba Experimental

5.1. Datos para el entrenamiento

De una población de 21 series temporales se seleccionaron 10 series aleatoriamente, corresponden a observaciones entre los periodos enero de 2001 a septiembre de 2006, la información fue obtenida del *Instituto Nacional de Estadística INE* y corresponden a las serie temporales

- Índice de precios al consumidor IPC
- Tipo de cambio TC
- Exportaciones
- Importaciones
- Deuda externa
- Emisión monetaria
- Medio circulante
- Liquidez total
- Tasa interés efectiva
- Unidad de fomento a la vivienda UFV

Es importante denotar que se realizarán sesgos a los datos de prueba, es decir se crea un conjunto de las series temporales estudiadas cada uno con 21 observaciones entre los periodos 2005 a 2006, y un otro conjunto de series temporales con un promedio de 60 observaciones entre los periodos 2001 a 2006.

A cada conjunto se aplicarán los modelos ARIMA y redes neuronales con el objetivo de hallar el comportamiento de los modelos cuando el tamaño de las muestra tiende a crecer, estos procedimientos se realizan en la fase de pruebas.

5.2. Pruebas

La fase de pruebas se realiza sobre un registro 10 series temporales correspondiente a distintos eventos expuestos en el acápite de presentación de datos.

El proceso de pruebas consiste en obtener el modelo generador de la serie temporal utilizando la red neuronal, análogamente se obtiene un otro modelo utilizando la metodología ARIMA.

Modelos en base a Redes Neuronales

Una previa fase de experimental se llevo a la conclusión que la dimensión optima del vector de entrenamiento es R^3 es decir $X = (x_i, x_{i+1}, x_{i+2})$ cuyo valor deseado será la componente x_{i+4} en la sucesión de series temporales.

Otros resultados óptimos son los relacionados con la red neuronal relativos al número máximo de iteraciones (epochs=3000), error mínimo deseado (goal=0.001), factor de aprendizaje (lr=0.25), además con una arquitectura de entre dos y tres capas se obtiene buenos resultados.

Bajo estos criterios se realizan una serie de entrenamientos variando el número de neuronas y capas para las 10 series temporales estudiadas, los cuales se observan en los siguientes cuadros 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10.

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	0.5762	0.6812	0.8505	0.4941	0.4674	0.3503
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	0.8432	0.4406	0.3053	0.3400	0.1487	0.1547

Cuadro 5.1: Índice de precios al consumidor IPC

Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	4.633e-6	4.243e-6	3.303e-6	2.123e-6	3.252e-6	1.851e-6
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	1.418e-6	1.588e-6	1.630e-6	1.489e-6	1.207e-6	3.422e-6

Cuadro 5.2: Tipo de cambio TC

Fuente: Elaboración propia

Los cuadros proveen información respecto a la arquitectura óptima aplicada a cada serie temporal estudiada, utilizando como criterio de selección el error cuadrático medio se obtiene el cuadro 5.11 que muestra una arquitectura específica a ser aplicada a cada serie temporal.

A las series temporales en estudio se aplicó las curvas de ajuste de los modelos de redes neuronales y ARIMA, los cuales se observan en las figuras 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	197.4954	193.8013	54.9144	96.9827	180.0259	316.8073
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	115.6449	26.0333	76.7908	95.4071	128.2210	18.7657

Cuadro 5.3: Exportaciones

Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	150.2361	8.7870	1.2247	3.0159	0.6660	71.4353
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	8.7721	0.6646	0.6658	0.6649	0.6661	0.6640

Cuadro 5.4: Importaciones
Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	8.9395e4	1.0781e4	5.5092e3	3.5027e3	1.2174e5	4.1721e3
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	2.7178e3	4.4270e3	2.7214e3	4.0786e4	8.7449e4	5.0254e3

Cuadro 5.5: Deuda Externa
Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	5.8916e4	2.5931e4	2.0322e4	6.3598e3	2.8122e3	3.6276e3
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	1.0694e4	1.4160e4	3.3608e3	1.8187e4	7.2708e3	1.0823e3

Cuadro 5.6: Emisión monetaria
Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	7.0167e4	2.4567e4	6.3615e4	3.2950e5	1.3752e4	8.8786e3
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	3.7619e4	4.6764e3	4.5514e3	8.6651e3	8.8265e3	4.2867e4

Cuadro 5.7: Medio circulante
Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	1.2152e5	7.4204e4	5.2100e4	3.5982e5	2.9874e4	3.1537e4
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	3.4194e4	4.0363e4	5.9537e4	1.8520e4	2.2208e4	2.7551e4

Cuadro 5.8: Liquidez total
Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	1.216e-4	0.0154	4.127e-4	4.893e-4	0.0106	0.1335
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	0.0212	4.124e-4	0.0022	4.131e-4	4.123e-4	0.0024

Cuadro 5.9: tasa de interés efectiva en moneda extranjera
Fuente: Elaboración propia

Capa 1	3	5	7	9	15	17
Capa 2	1	1	1	1	1	1
ECM	5.605e-7	5.609e-7	5.560e-7	5.615e-7	5.616e-7	1.113e-6
Capa 1	7	9	9	15	20	20
Capa 2	3	5	7	10	10	15
Capa 3	1	1	1	1	1	1
ECM	5.614e-7	5.613e-7	5.612e-7	5.605e-7	5.600e-7	5.615e-7

Cuadro 5.10: Unidad de fomento a la vivienda UFV
Fuente: Elaboración propia

Serie Temporal	Arquitectura RNA
Indice de precios IPC	[20 10 1]
Tipo de cambio TC	[20 10 1]
Exportaciones	[9 5 1]
Importaciones	[20 15 1]
Deuda externa	[7 3 1]
Emisión monetaria	[20 15 1]
Medio circulante	[9 7 1]
Liquidez total	[15 10 1]
Tasa interés	[9 5 1]
U.F. vivienda UFV	[17 1]

Cuadro 5.11: Modelo ARIMA asociado a la serie temporal
Fuente: Elaboración propia

MES	INDICE	
	2005	2006
Enero	236.97	235.84
Febrero	227.09	236.58
Marzo	227.46	235.92
Abril	236.59	236.25
Mayo	238.12	238.23
Junio	231.63	239.71
Julio	230.56	241.06
Agosto	231.37	241.51
Septiembre	231.71	241.71
Octubre	232.58	
Noviembre	233.59	
Diciembre	234.89	

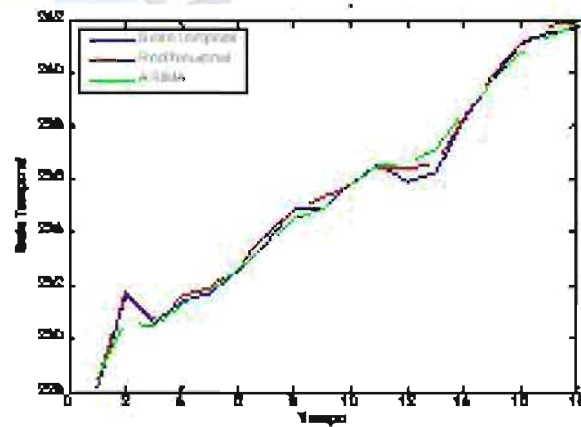


Figura 5.1: Índice de precios al consumidor IPC 2005 - 2006
Fuente: Elaboración propia.

Modelos ARIMA

Observando el comportamiento de las series estudiadas respecto a la estacionariedad, funciones de autocorrelación FAC y FACP se obtienen los modelos generadores ARIMA de las 10 series estudiadas, los cuales se observan en los siguientes cuadros 5.12, 5.13

5.3. Resultados

De el proceso experimental se denotan los siguientes resultados sobresalientes

- Los modelos de redes neuronales son óptimos cuando la longitud de la serie temporal

MES	TIPO DE CAMBIO	
	2005	2006
Enero	8,0655	8,0800
Febrero	8,0743	8,0736
Marzo	8,0819	8,0700
Abril	8,1000	8,0690
Mayo	8,1000	8,0600
Junio	8,1000	8,0600
Julio	8,0950	8,0600
Agosto	8,0900	8,0558
Septiembre	8,0840	8,0500
Octubre	8,0800	
Noviembre	8,0800	
Diciembre	8,0800	

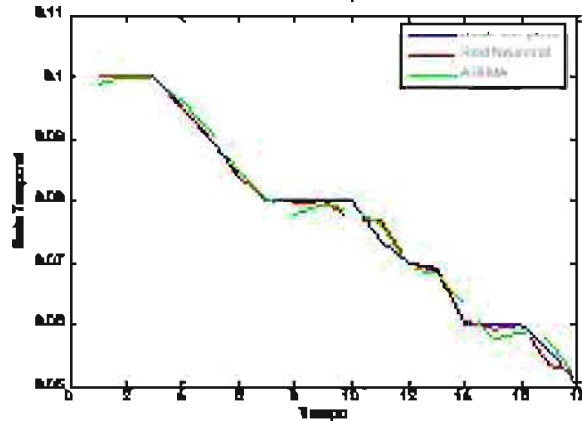


Figura 5.2: Tipo de cambio TC 2005 - 2006
Fuente: Elaboración propia.

MES	MENSUAL	
	2005	2006
Enero	177,39	280,38
Febrero	182,00	279,02
Marzo	207,21	304,51
Abril	225,82	306,32
Mayo	254,21	341,54
Junio	213,15	353,89
Julio	256,26	374,79
Agosto	280,40	388,94
Septiembre	239,79	349,02
Octubre	279,02	
Noviembre	256,31	
Diciembre	241,25	

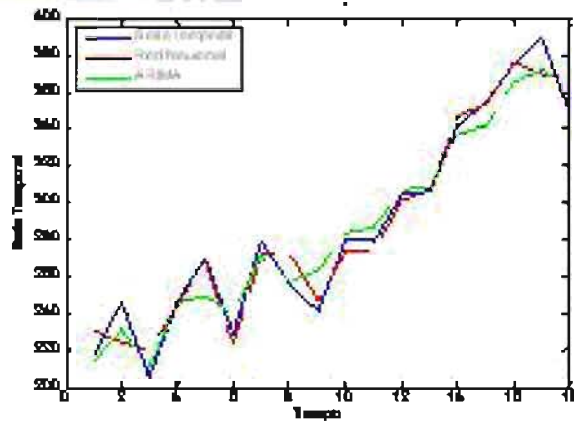


Figura 5.3: Exportaciones 2005 - 2006
Fuente: Elaboración propia.

MES	MENSUAL	
	2005	2006
Enero	171,18	228,38
Febrero	162,69	175,14
Marzo	161,10	257,67
Abril	189,13	224,34
Mayo	187,17	272,90
Junio	186,18	244,81
Julio	206,45	206,85
Agosto	211,87	222,38
Septiembre	201,49	193,50
Octubre	212,81	
Noviembre	229,19	
Diciembre	224,04	

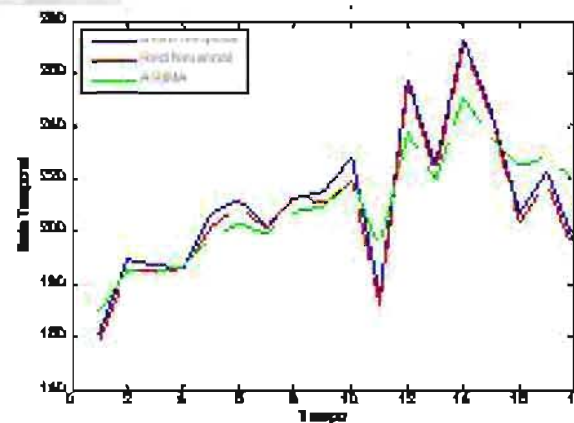


Figura 5.4: Importaciones 2005 - 2006
Fuente: Elaboración propia.

MES	DEJDA	
	2005	2006
Enero	4.883,40	4.729,74
Febrero	4.868,60	4.661,99
Marzo	4.867,63	4.663,59
Abril	4.987,00	4.689,48
Mayo	4.821,40	4.709,65
Junio	4.802,00	4.673,30
Julio	4.787,80	3.191,49
Agosto	4.828,80	3.203,20
Septiembre	4.807,80	3.206,48
Octubre	4.800,70	
Noviembre	4.767,60	
Diciembre	4.941,60	

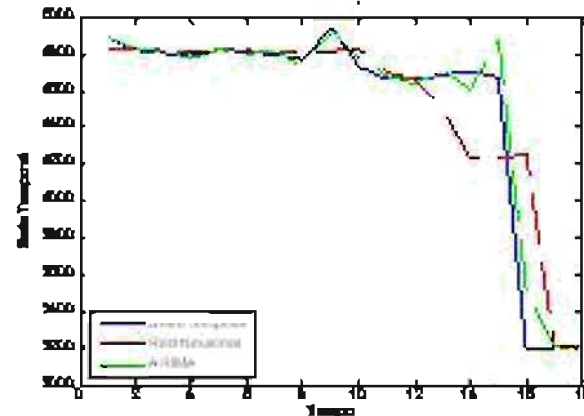


Figura 5.5: Deuda externa 2005 - 2006

Fuente: Elaboración propia.

MES	EMISION	
	2005	2006
Enero	3.751,69	5.649,94
Febrero	3.692,72	5.668,06
Marzo	3.571,87	5.497,84
Abril	3.898,22	5.795,83
Mayo	3.913,73	6.023,95
Junio	4.044,46	6.392,12
Julio	4.165,15	6.677,42
Agosto	4.365,18	6.718,05
Septiembre	4.577,93	6.997,97
Octubre	4.767,23	
Noviembre	4.888,49	
Diciembre	6.179,51	

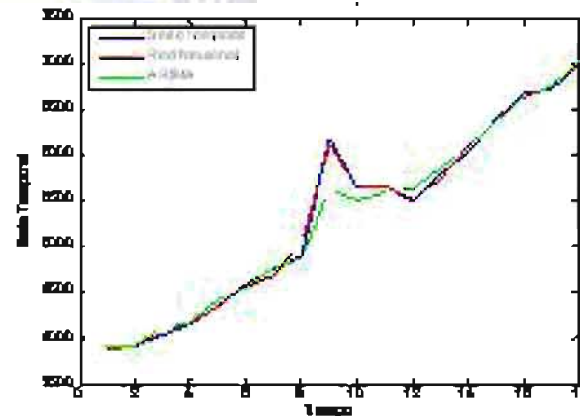


Figura 5.6: Emisión monetaria 2005 - 2006

Fuente: Elaboración propia.

MES	CIRCULANTE	
	2005	2006
Enero	8.770,92	10.958,18
Febrero	8.592,82	11.070,40
Marzo	8.477,04	10.791,15
Abril	8.857,09	11.047,28
Mayo	8.906,72	11.123,39
Junio	8.891,05	11.435,95
Julio	9.178,57	11.967,14
Agosto	9.753,10	12.245,09
Septiembre	9.937,70	12.603,45
Octubre	10.712,12	
Noviembre	11.667,95	
Diciembre	11.483,33	

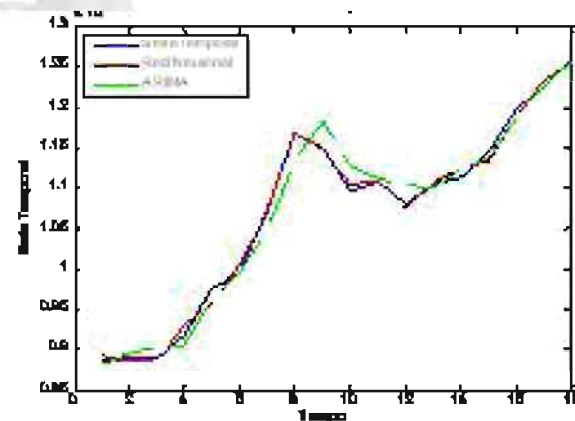


Figura 5.7: Medio circulante 2005 - 2006

Fuente: Elaboración propia.

MES	LIQUIDEZ	
	2005	2006
Enero	29.706,89	33.492,52
Febrero	29.806,77	33.696,00
Marzo	29.615,73	33.663,91
Abril	30.691,11	33.970,14
Mayo	31.180,84	33.904,95
Junio	31.016,67	34.228,86
Julio	31.520,95	34.697,31
Agosto	32.446,92	35.498,36
Septiembre	32.830,21	36.095,64
Octubre	33.575,72	
Noviembre	34.332,99	
Diciembre	34.312,95	

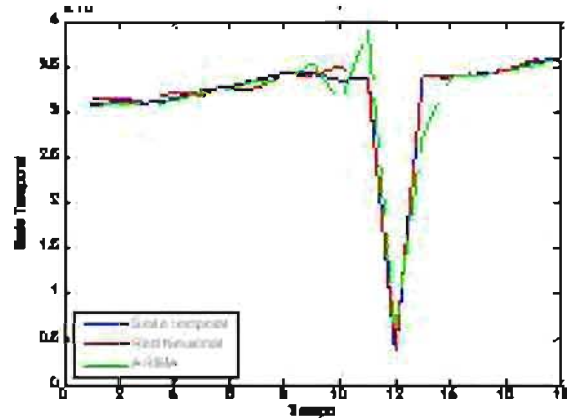


Figura 5.8: Liquidez total 2005 - 2006
Fuente: Elaboración propia.

MES	ACTIVAS	
	2005	2006
Enero	11,54	11,49
Febrero	11,44	11,87
Marzo	11,16	11,30
Abril	11,45	11,11
Mayo	10,29	11,21
Junio	10,20	12,30
Julio	11,12	10,90
Agosto	11,47	
Septiembre	11,93	
Octubre	11,63	
Noviembre	12,10	
Diciembre	11,60	

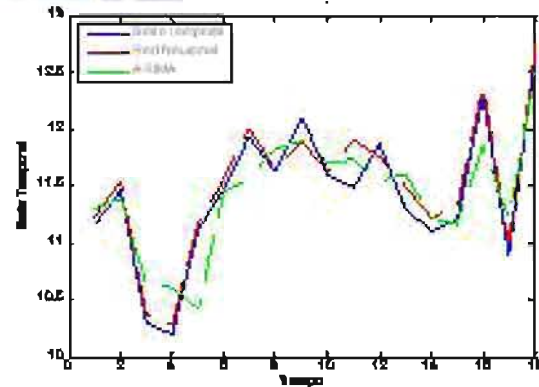


Figura 5.9: Tasa interés efectiva 2005 - 2006
Fuente: Elaboración propia.

MES	VALOR	
	2005	2006
Enero	1,09008	1,14887
Febrero	1,09473	1,15275
Marzo	1,09944	1,15664
Abril	1,10446	1,16025
Mayo	1,10928	1,16420
Junio	1,11425	1,16838
Julio	1,11969	1,17189
Agosto	1,12494	1,17606
Septiembre	1,12984	1,18030
Octubre	1,13489	
Noviembre	1,13956	
Diciembre	1,14429	

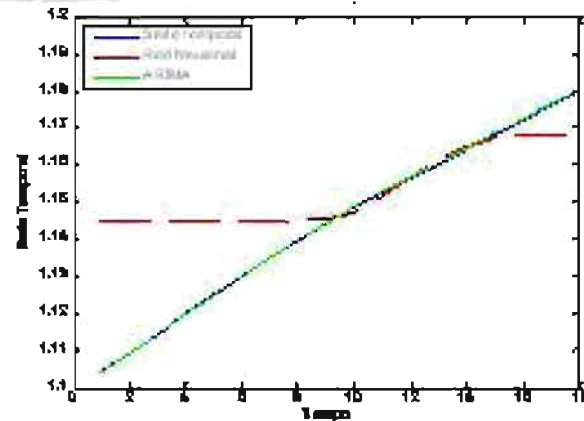


Figura 5.10: Unidad de fomento a la vivienda UFV 2005 - 2006
Fuente: Elaboración propia.

Serie Temporal	Modelo ARIMA
Indice de precios IPC	$y_t = 1,003501y_{t-1} + \varepsilon_t + 0,996891\varepsilon_{t-1}$
Tipo de cambio TC	$y_t = 0,999833y_{t-1} + \varepsilon_t - 0,932328\varepsilon_{t-1}$
Exportaciones	$y_t = 1,035079y_{t-1} + \varepsilon_t + 0,092128\varepsilon_{t-1}$
Importaciones	$y_t = 0,994303y_{t-1} + \varepsilon_t$
Deuda externa	$y_t = 0,982110y_{t-1} + \varepsilon_t - 0,927894\varepsilon_{t-1}$
Emisión monetaria	$y_t = 1,032327y_{t-1} + \varepsilon_t$
Medio circulante	$y_t = 1,024121y_{t-1} + \varepsilon_t - 0,9972001\varepsilon_{t-1}$
Liquidez total	$y_t = 0,991735y_{t-1} + \varepsilon_t - 2,258050\varepsilon_{t-1}$
Tasa interés	$y_t = 1,000049y_{t-1} + \varepsilon_t - 0,643079\varepsilon_{t-1}$
U.F. vivienda UFV	$y_t = 1,941577y_{t-1} + \varepsilon_t + 0,941368\varepsilon_{t-1}$

Cuadro 5.12: Modelo ARIMA asociado a la serie temporal
Fuente: Elaboración propia

Serie Temporal	Modelo	BIC	AIC
Indice de precios IPC	ARIMA(1,1,1)	2.397586	2.792409
Tipo de cambio TC	ARIMA(1,1,1)	10.14710	9.997981
Exportaciones	ARIMA(1,1,1)	8.568490	8.419369
Importaciones	AR(1)	9.803630	9.753843
Deuda externa	ARIMA(1,1,1)	13.46562	13.31649
Emisión monetaria	AR(1)	13.69175	13.59233
Medio circulante	ARIMA(1,1,1)	13.66005	13.51092
Liquidez total	ARIMA(1,1,1)	18.65827	18.50915
Tasa interés	ARIMA(1,1,1)	0.879444	.731048
U.F. vivienda UFV	ARIMA(1,1,1)	-13.48349	13.63188

Cuadro 5.13: Criterios de informacion AIC BIC
Fuente: Elaboración propia

es pequeña, alrededor de 21 observaciones, si se incrementa este tamaño en una cota mínima de 60 observaciones los errores cuadráticos medios tienden a crecer.

- Los modelos ARIMA presentan mayor estabilidad al incrementarse la longitud de la serie temporal, los errores cuadráticos medios decrecen al contrario que una red dichos errores tienden a crecer, este comportamiento se puede observar en los cuadros 5.18, 5.19.
- Se trabajó con funciones de activación tangente sigmoideal y logaritmo sigmoideal, este último presenta mayor estabilidad fue utilizada en las capas ocultas, en la capa de salida la función identidad y responde al entrenamiento de todas las series estudiadas.

5.4. Pesos para índice de exportaciones

A efecto de observar la aproximación numérica de los modelos RNA y ARIMA, la distribución de pesos en las capas ocultas y de salida de acuerdo a su arquitectura [9 5 1] se seleccionará al índice de exportaciones, en los cuadros 5.14, 5.15, 5.16 se observan los pesos correspondiente a las capas 1 2 3, de forma análoga el cuadro 5.17 muestra los valores numéricos de aproximación, el error cuadrático medio en el proceso de ajuste.

Neurona 1	1.7058	-1.8231	-1.7512
Neurona 2	1.3796	3.3465	0.1013
Neurona 3	2.8862	2.4301	-3.8786
Neurona 4	0.9896	-2.2515	1.4379
Neurona 5	0.7546	-4.1185	-1.8453
Neurona 6	3.5866	2.7158	-1.0863
Neurona 7	-4.3349	-0.3840	-0.5298
Neurona 8	2.4292	-1.6939	-1.5798
Neurona 9	-2.4596	1.8685	-0.7152

Cuadro 5.14: Distribución pesos primera capa oculta
Fuente: Elaboración propia

Neurona 1	Neurona 2	Neurona 3	Neurona 4	Neurona 5
-2.3649	0.5924	1.9873	0.3070	0.9856
-2.4869	-0.9841	2.1464	1.6633	-1.6146
0.5020	4.6634	-3.3574	0.9191	-4.6969
-0.7071	-1.6624	2.5805	-4.4506	1.9407
4.0609	5.9733	1.9814	-0.9393	0.0879
-2.4269	-2.4269	-1.7994	4.1498	2.4546
-0.7602	-0.7602	-1.6605	-0.8706	3.2338
1.6293	1.6293	2.1239	0.7030	-3.0205
-2.5275	-2.5275	1.0407	-1.1504	0.5940

Cuadro 5.15: Distribución pesos en la segunda capa oculta
Fuente: Elaboración propia

Neurona 1	1.3982	-2.7734	0.5394	-0.2547	-1.6384
-----------	--------	---------	--------	---------	---------

Cuadro 5.16: Distribución pesos en la capa de salida
Fuente: Elaboración propia

Exportación	Modelo RNA	Modelo ARIMA
225.82	225.5083	209.724
254.21	224.3764	237.193
213.15	225.8321	223.479
256.26	248.7686	267.857
280.40	273.0732	255.246
239.79	234.7663	252.459
279.02	272.0879	293.298
256.31	248.7903	255.896
241.25	238.3977	269.829
280.38	275.6034	281.925
279.02	274.3582	267.728
304.51	295.2385	302.308
306.32	300.3977	301.705
341.54	335.1506	334.667
353.89	349.3938	339.857
374.79	371.0147	369.399
388.94	372.6256	383.396
349.02	349.5405	372.576
ECM	26.0333	201.0415

Cuadro 5.17: Contraste de resultados modelos RNA ARIMA

Fuente: Elaboración propia

5.5. Comparación de resultados

En el proceso experimental se obtuvieron los modelos RNA y ARIMA correspondientes a las series temporales estudiadas, en el cuadro 5.18 se presenta un resumen respecto a los errores cuadráticos medios para una longitud de 21 observaciones.

En el cuadro 5.19 se observa un cambio favorable de los errores respecto a los modelos ARIMA cuando la longitud de la serie temporal se incremento a 60 observaciones como cota mínima.

5.6. Discusión de resultados

De una población muestreada de 10 series temporales que forman el 100% de las series estudiadas, un 20% corresponde a los modelos ARIMA y el 80% a los modelos RNA, esto implica que los modelos basados en redes neuronales artificiales realizan un mejor

Serie Temporal	RNA ECM	ARIMA ECM	Optimo
Indice de precios IPC	0.0592	0.1660	RNA
Tipo de cambio TC	1.7911e-006	2.1679e-006	RNA
Exportaciones	116.2491	120.8856	RNA
Importaciones	14.0351	155.2233	RNA
Deuda externa	9.0301e+004	8.8373e+003	ARIMA
Emisión monetaria	2.3600e+003	2.0355e+004	RNA
Medio circulante	6.8528e+003	2.9193e+004	RNA
Liquidez total	2.4911e+005	4.9214e+006	RNA
Tasa interés	0.0201	0.0901	RNA
U. F. vivienda UFV	2.9553e-004	6.3550e-008	ARIMA

Cuadro 5.18: Modelo óptimo criterio ECM 21 observaciones
Fuente: Elaboración propia

Serie Temporal	RNA ECM	ARIMA ECM	Optimo
Indice de precios IPC	1.0404	0.2229	ARIMA
Tipo de cambio TC	0.0486	0.0673	RNA
Exportaciones	225.9229	113.3372	ARIMA
Importaciones	160.1750	116.0444	ARIMA
Deuda externa	1.7406e+005	6.7302e+004	ARIMA
Emisión monetaria	4.5152e+004	2.7680e+004	ARIMA
Medio circulante	1.0629e+005	3.1690e+004	ARIMA
Liquidez total	1.0859e+006	6.4768e+006	RNA
Tasa interés	0.2481	0.1363	ARIMA
U. F. vivienda UFV	3.3285e-006	4.7641e-007	ARIMA

Cuadro 5.19: Modelo óptimo criterio ECM 60 observaciones
Fuente: Elaboración propia

proceso de ajuste a las series temporales estudiadas, cuando la longitud de dichas series es relativamente pequeña, alrededor de 21 observaciones.

Si el tamaño de la serie temporal supera las 60 observaciones la efectividad de los modelos ARIMA tiende a crecer, se observa en el cuadro 5.19 que de un 100% de las series temporales un 80% corresponde a los modelos ARIMA y un 20% a los modelos RNA

Estos resultados se deben a que la red neuronal al realizar gran cantidad de cálculos numéricos pierde estabilidad debido a que se trabaja con grandes volúmenes de información.

5.7. Aplicación en Medicina

La aplicación de la red neuronal sobre series temporales permitió hallar el modelo generador de la serie estudiada el cual puede ser utilizado en la medicina para el reconocimiento de ondas provenientes de electrocardiogramas cuya característica principal son los complejos P, Q, R, S, T que interpretan los latidos del corazón, ver Figura 5.11

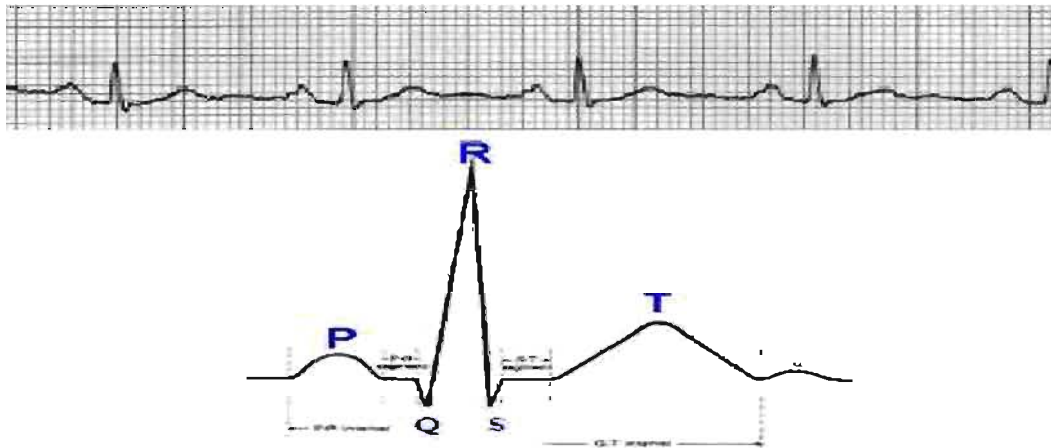


Figura 5.11: Electrocardiograma Normal Ondas P, Q, R, S, T

Fuente: Elaboración propia

Pruebas

Se seleccionaron 20 electrocardiogramas correspondientes a la Fibrilación Ventricular que equivale clínicamente a un paro cardíaco, los cuales fueron analizados por la red neuronal originando los siguientes resultados

- Del 100% de los electrocardiogramas analizados por la red neuronal un 90% son reconocidos.
- La similitud de los electrocardiogramas de prueba con los patrones de entrenamiento es un factor decisivo para que la red aprenda a reconocer este nuevo patrón.
- La red neuronal no reconoce un ECG de prueba cuando presenta comportamientos bastante difusos respecto de los patrones de entrenamiento.
- El índice de reconocimiento puede ser mejorado al incluir al conjunto de patrones de entrenamiento una gama mayor de electrocardiogramas.

Capítulo 6

Conclusiones y Recomendaciones

6.1. Conclusiones

Luego de un proceso de investigación se cuenta con varios resultados los cuales no siempre pueden ser replicados a exactitud debido a la naturaleza de su procedencia, tal es el caso de las redes neuronales artificiales que trabajan con factores aleatorios.

Relativas a las Series Temporales

- a) Una serie temporal que contiene 60 o más observaciones es óptimo para su estudio, los modelos ARIMA se ajustan con mejor precisión que un modelo de redes neuronales.
- b) La longitud de la series temporal con un número de 21 observaciones promedio, los modelos de redes neuronales son óptimos, se ajustan con mayor precisión que un modelo ARIMA.

Relativas a las Redes Neuronales

- a) El modelo de redes neuronales artificiales “*Backpropagation*” presenta cierta inestabilidad cuando la cantidad de información a ser tratada tienda a crecer significativamente ($n > 60$).

- b) Los parámetros: Número de iteraciones (*show*), tasa de aprendizaje (*lr*), error mínimo (*goal*), no son posibles de generalizar a otros ejemplos de series temporales, se adaptan a nuevos requerimientos en un proceso de prueba y error.
- c) En la creación de la RNA los pesos se inicializan aleatoriamente debido a ello no siempre convergen a un único resultado, se deben realizar una secuencia de entrenamientos hasta hallar un resultado satisfactorio y grabar estos pesos.
- d) la función de activación logaritmo sigmoideal "*logsig*" es mas robusto y estable soporta el incremento del número de capas y neuronas lo que no ocurre con la función tangente sigmoideal "*tansig*" que presenta inestabilidad en el entrenamiento, al incrementarse el numero de capas y neuronas.
- e) Preparar una red neuronal para alguna aplicación específica es más un arte que proceso metodológico estructurado debido a ello no se le puede atribuir un nivel alto de relevancia en la solución de algún problema, a su favor esta la tolerancia fallos en los datos y procesos.

Relativas a los modelos ARIMA

- a) Los modelos ARIMA están diseñados precisamente para el trabajo con series temporales, el nivel de confiabilidad es alentador al manejar grandes cantidades de información, por tanto es una herramienta muy útil para un analista de series temporales.
- b) Aplicar el proceso ARIMA requiere conocer los fundamentos matemáticos que permitan precisar el modelo generador de la serie temporal, esta ligado a la experiencia y destreza del analista.
- c) Los factores de retardo p , q , el nivel de integración, las funciones de auto correlación FAC, FACP, los criterios de información AIC, BIC son vitales en el proceso de identificación del modelo generador de la serie temporal.
- d) ARIMA es una metodología no ligado a los factores aleatorios razón por el que luego de un proceso de búsqueda se replica nuevamente el modelo anteriormente obtenido.

6.2. Recomendaciones

Relativa a las Series Temporales

- a) Es vital conocer los principios de estacionariedad, los momentos de primer, segundo y tercer orden que dan origen a la esperanza, varianza y covarianza respectivamente, los que contribuirán a tener un idea clara de la temática de investigación.
- b) Orientar el proceso de investigación en sentido práctico en las áreas biológicas, medio ambiente, finanzas, empresariales, medicina los cuales contribuirán a la solución de problemas reales.

Relativa a las Redes Neuronales Artificiales

- a) Incursionar en otros tipos de redes tales como: Las redes de Elman, Hopfield, Estocásticas, redes de Base Radial, será interesante observar los resultados que proveerán su aplicabilidad en los procesos estocásticos.
- b) En el presente trabajo de investigación se utilizo criterios de *heteroasociación* de la información de *n-sucesiones anteriores a una siguiente* entonces será interesante deducir otro tipo de criterios al respecto.

Relativa a los modelos ARIMA

- a) Para todo analista de Series Temporales es vital conocer los fundamentos matemáticos que originan el proceso ARIMA ello implica la exploración de teorías, conceptos que pueden ser abordados con el asesoramiento de expertos en el área.
- b) El software utilizado en el proceso ARIMA fue el Lenguaje R 2.1.1, Eviews 5.0 será de interés incursionar en otras herramientas tales como SPS, SAS.
- c) El trabajo con los modelos ARIMA en el presente trabajo, habré horizontes a nuevas aplicaciones tales como la medicina en la detección de patologías.

Una Serie Temporal es un evento que no necesariamente tiene que estar asociado a la actividad humana si no a la naturaleza misma que nos provee un inicio en el entendimiento de la complejidad que nos rodea.

La búsqueda constante de nuevas herramientas en distintas áreas del conocimiento humano son un reto a alcanzar.

Noviembre 2006.



Bibliografía

- [1] DIEPENBEEK, L. and LIEGE, U. (2001): *The Statistical Analysis of Stochastic Processes in Time*, United States of America 2001.
- [2] URIEL, J. EZEQUIEL Y PEIRO, G. AMADO. (2000): *Introducción al Análisis de Series Temporales*, Editorial AC, Madrid España 2000.
- [3] EMANUEL, PARZEN. (1972): *Procesos Estocásticos*, Editorial Holden Day. INC, San Francisco USA 1972.
- [4] NILS, J. NILSSON. (2001): *Inteligencia Artificial una Nueva Síntesis*, Editorial Mac Graw Hill, Aravaca Madrid 2001.
- [5] HILERA, R. JOSE y MARTÍNEZ, J. VICTOR (2000): *Redes Neuronales Artificiales*, Editorial Ra-ma, Madrid España 2000.
- [6] RUSSEL, STUART Y NORVIG, PETER. (1996): *Inteligencia Artificial un Enfoque Moderno*, Editorial Prentice Hall Hispanoamericana S.A. Naucalpan de Juárez México 1996.
- [7] RICH, ELIANE Y KNIGHT, KEVIN. (1994): *Inteligencia Artificial*, Editorial Mc Graw Hill, Aravaca Madrid 1994.
- [8] HUNT, V. DANIEL (1986): *Artificial Intelligence and Expert Systems Sourcebook*, Editorial Chapman and Hall, United States of America 1986.
- [9] FLORES, J. CARLOS. (1997): *Modelos Estructurales y Aplicaciones*, Tesis de Grado Carrera Estadística, UMSA, La Paz Bolivia 1997.
- [10] MARQUIEGUI, N. RAUL (1994): *Procesos Estocásticos*, Tesis de Grado Carrera Estadística, UMSA, La Paz Bolivia 1994.
- [11] CHAVEZ, NICOLÁS. (1991): *aplicación de los modelos ARIMA a la econometria*, Tesis de Grado Carrera Estadística, UMSA, La Paz Bolivia 1991.
- [12] CUARITA, A. GABRIELA. (2000): *Predicción y Simulación Estocástica de Series Temporales*, Tesis de Grado Carrera Informática, UMSA, La Paz Bolivia 2000.
- [13] ARISTA, M. ROSALIA. (1998): *Redes Neuronales en la Predicción de series de Tiempo*, Tesis de Grado Carrera Informática, UMSA, La Paz Bolivia 1998.